TECHNIQUES FOR ESTIMATING THE MAGNITUDE AND FREQUENCY OF FLOODS IN MINNESOTA

U.S. GEOLOGICAL SURVEY
Water-Resources Investigations Report 87-4170

Prepared in cooperation with the
MINNESOTA DEPARTMENT OF TRANSPORTATION

TECHNIQUES FOR ESTIMATING THE MAGNITUDE AND FREQUENCY OF FLOODS IN MINNESOTA

By J. E. Jacques and D. L. Lorenz

U.S. GEOLOGICAL SURVEY

Water-Resources Investigations Report 87-4170

Prepared in cooperation with the

MINNESOTA DEPARTMENT OF TRANSPORTATION

St. Paul, Minnesota

1988

CONVERSION FACTORS AND ABBREVIATIONS

Readers who prefer to use metric (International System) units rather than inch-pound units can make conversions using the following factors:

Multiply Inch-Pound Unit	By	To obtain Metric Unit
foot (ft)	0.3048	meter (m)
mile (mi)	1.609	kilometer (km)
square mile (mi^2)	2.590	square kilometer (km^2)
cubic foot per second (ft^3/s)	.02832	cubic meter per second (m^3/s)
foot per mile (ft/mi)	.1894	meter per kilometer (m/km)

TECHNIQUES FOR ESTIMATING THE MAGNITUDE AND FREQUENCY OF FLOODS IN MINNESOTA

By J. E. Jacques and D. L. Lorenz

ABSTRACT

Log-Pearson type III flood-frequency analyses were made of annual series peak-flow records from 246 gaging stations on unregulated streams in Minnesota having watersheds ranging in area from 0.08 to 2,520 square miles. These flood discharges were related to watershed and climatic characteristics by using multiple-regression techniques. On the basis of this preliminary regression analysis of the frequency-analysis results, the data from these stations were grouped into four hydrologically distinct regions for the State. Regression analyses were performed on data from each region relating the 2-, 5-, 10-, 25-, 50-, and 100-year recurrence interval flood discharges to basin characteristics. The resulting regression equations, which may be used to estimate flood flows at ungaged sites, relate basin characteristics (contributing drainage area, main-channel slope, percent of basin covered by water, percent of basin covered by lakes, and mean annual runoff) to estimated flood flows. Different basin characteristics are significant for each of the four regions. Drainage area was found to be most significant and is included in all regional equations. Standard errors of estimate of the regression equations ranged from 33 to 60 percent.

INTRODUCTION

Background

Knowledge of the magnitude and frequency of floods is essential for regulation, planning, and design along Minnesota's rivers and streams. Ideally, discharge information necessary for such projects would be obtained by hydrologic analysis of nearby long-term flood records from gaging stations on the rivers and streams. Because such records are rarely available at all sites of interest, particularly in small basins, techniques are needed to estimate the magnitude and frequency of floods at ungaged sites.

This report is one of a series of reports prepared in cooperation with the Minnesota Department of Transportation that discuss flood-flow-frequency on small streams. The U.S. Water Resources Council (1981a) indicates that regression equations present a more accurate technique than other methods tested (for example the rational equation and rainfall-runoff models).

Annual maximum-discharge and basin-characteristic data used in this study are stored, maintained, and updated by the U.S. Geological Survey in the National Water Data Storage and Retrieval System (WATSTORE). In addition, these data are published by the U.S. Geological Survey as part of cooperative programs with various local, State, and Federal agencies. For this study, gaging stations with less than 10 years of record were not used because of the increased probability of time-sampling errors. Stations with more than 3,000 square miles of drainage area and stations influenced by natural and (or) manmade regulation also were excluded. Application of these criteria resulted in the selection of 246 gaging stations in Minnesota (57 continuous record, 140 crest stage, and 49 combination sites).

Purpose and Scope

The purposes of this report are to (1) describe the analytical techniques used for annual series flood-frequency computations, regionalization, and development of estimating equations for small watersheds, (2) present flood-frequency data at gaged sites, and (3) develop equations for estimating flood-flows and present examples of flood-flow estimations at gaged and ungaged sites on unregulated streams.

This report supersedes previous reports by Prior (1949), Prior and Hess (1961), Wiitala (1965), Patterson and Gamble (1968), and Guetzkow (1977); all of which dealt with techniques for estimating flood magnitudes in Minnesota.

ANALYTICAL TECHNIQUES

Flood-Frequency Analysis of Gaging Stations

An annual series peak-flood-frequency analysis at each gaging station was prepared according to the procedures outlined in Bulletin 17B [U.S. Water Resources Council, 1981b.] Federal agencies are requested to use these guidelines for all analysis of flood frequencies of unregulated streams. The equation for fitting the log-Pearson Type III frequency-distribution function to the T-year recurrence interval is defined below:

$$\text{Log } Q_T = M + KS$$

where:

Q_T is the peak discharge for T-year recurrence interval,

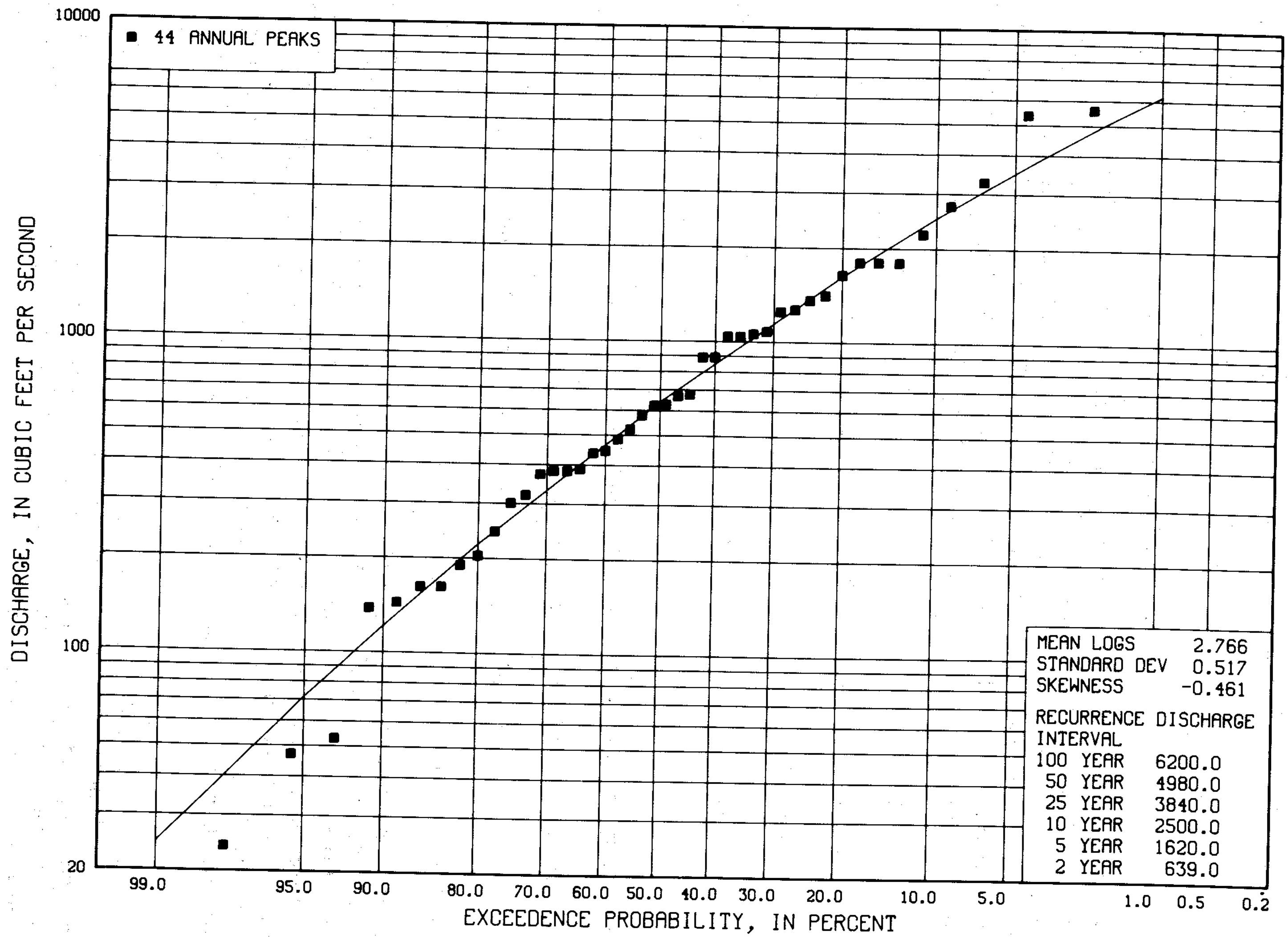
M is the mean of the logarithms of annual peaks,

K is a factor dependent on T and the coefficient of skewness available from Bulletin 17B, appendix 3, and

S is the standard deviation of the logarithms of annual peaks.

An example plot from the Log-Pearson analysis of the Redwood River near Marshall, Minnesota, (station number 05315000) is shown on figure 1.

Recorded flood peaks at some stations with less than 25 years of record exceed the high-outlier criterion defined in Bulletin 17B. Stations in an area in which a well-documented historic flood occurred were considered for adjustment in the peak-flow analysis. A regional analysis of the occurrence of the historic floods was used to evaluate whether a peak of record also should be included in the historic period. Criteria in Bulletin 17B for identifying low and high outliers were used in all cases.



Log-Pearson type III flood-frequency estimate for stream-gaging station

05315000 REDWOOD RIVER NEAR MARSHALL

Figure 1.--Example of flood-frequency plot.

Figure 2 shows the locations of the 246 gaging stations that were used in this analysis.

Multiple-Regression Analysis

Multiple-linear-regression techniques were used to define relations between flood flows and basin characteristics. Previous studies by Benson (1962) and Guetzkow (1977) have shown that the logarithmic transform of the data results in linear relations between flood flows and basin characteristics. This study used the logarithmic transformation that results in general linear-regression models described by the equations below.

For transformed variables:

$$\text{Log } Q_T = B_0 + B_1 \log X_1 + B_2 \log X_2 + \cdots + B_n \log X_n$$

or as untransformed variables:

$$Q_T = e^{B_0} (X_1)^{B_1} (X_2)^{B_2} \cdots (X_n)^{B_n}$$

where:

Q_T is the peak discharge for T-year recurrence interval,

B_i are regression coefficients,

X_i are independent variables (basin characteristics), and

e is the base of the natural logarithms.

The stepwise method of multiple-regression analysis used is described in Hocking (1976). Only those independent variables statistically significant at the 10-percent level of significance are included in the equations. The 10-percent criterion was used to standardize, on a regional basis, the basin characteristics used to define flood flows at the various recurrence intervals. The inclusion of the same variables in all equations for a region improves the continuity of the frequency curves constructed from the equations.

Bevington (1969, p. 100-102) and Draper and Smith (1981, p. 108) indicate that when the variance of the dependent variable is not constant for all observations, the equations resulting from the regression analysis may be a poor estimate of the "true" relation. Because it is well known that predicted flow magnitudes are more accurate from long-term gaging records than from short-term records (Linsley and others, 1982, p. 358), residuals were analyzed to determine any trend as a function of gaging-record length. A significant decrease in the variance of the residuals was observed as gaging record increased. Bobee (1973) defines an equation to estimate the variance of a calculated flood magnitude given the return period, standard deviation, coefficient of skewness, and the number of years of record. The variances of the flood estimates of different return periods were found to be proportional to the return period. Because of this relationship, only the calculated flood-magnitude variance for each station, for the 10-year return period, was

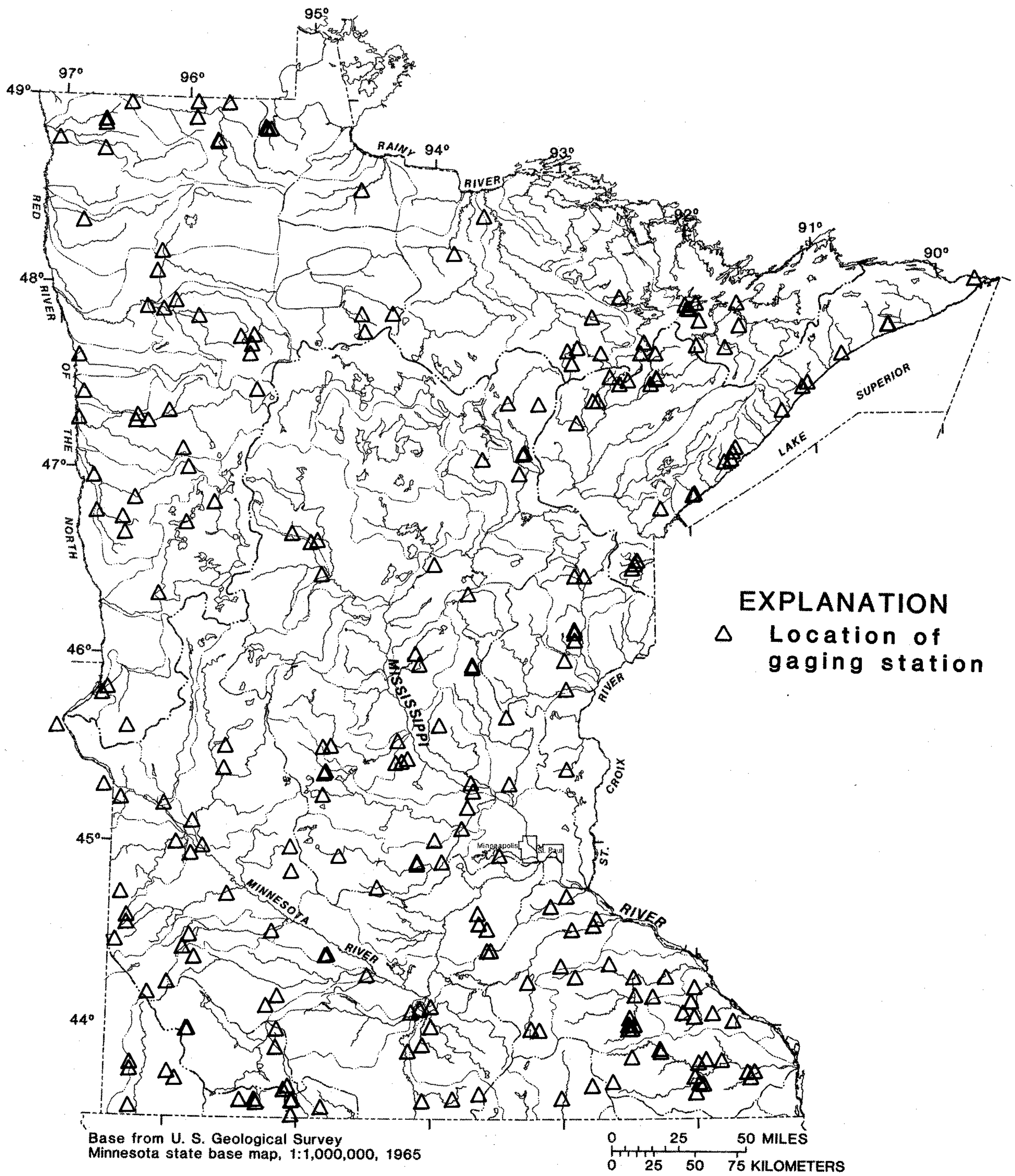


Figure 2.--Location of gaging stations used to define regression equations.

used to compute weights for the regression performed for this report. The resulting weighting factors for some stations deviate significantly from unity, apparently because of different basin characteristics and other unexplained factors that affect the standard deviation of the logarithms of the annual peaks. Thus, the reciprocal of the number of years of station record was normalized and averaged with the variance to define the final station weights;

$$S_i = (VAR(Q10_i) + a/N_i),$$

where $VAR(Q10_i)$ is the variance of the 10-year estimate, a is a normalizing factor, chosen to give equal weight to the variance and years of record, and N_i is the number of years of record for the station. The final station weight is the reciprocal of the sum of the variance and the reciprocal number of years of record;

$$W_i = b/S_i,$$

where W is the final station weight and b is chosen so that the sum of the weights equals the number of stations used in the regression analysis.

Regional Analysis

Hydrologic regions shown on figure 3 were defined by several techniques of regional analysis. Preliminary residual analysis indicated within the state five loosely defined regions. Subsequent analysis of residuals of equations obtained for each region indicated that only minor boundary adjustments were necessary to reduce regional bias that was not indicated by the analysis of clusters. The regression equations from two of the original regions were very similar and were joined into one group.

Intraregional comparisons of the regression equations shows differences which may help understand factors that affect flood magnitudes in each region. For example, the coefficients in the equations for region C are all similar except for slope. This implies that slope is the major determining factor, in region C, for the differences between streams for the various T-year flood magnitudes. Compare this to region A, where only the coefficient for area is roughly constant.

Equations derived by further subdivision of these regions did not vary significantly from those for the region as a whole. For example, region D was divided into 5 subregions and regression analysis performed on the data with different subregions deleted. The results from those regression analyses did not vary significantly from the regression on the whole dataset.

Regional boundaries outlined on figure 3 generally follow basin divides. The single exception is the boundary between regions B and C where the boundary crosses the watershed of the St. Louis River. Headwaters of the St. Louis River are in a flat region, but the tributaries to Lake Superior are very steep, similar to region C. Based on these topographic and geologic features, the boundary between regions B and C crosses the St. Louis River below Thompson Reservoir.

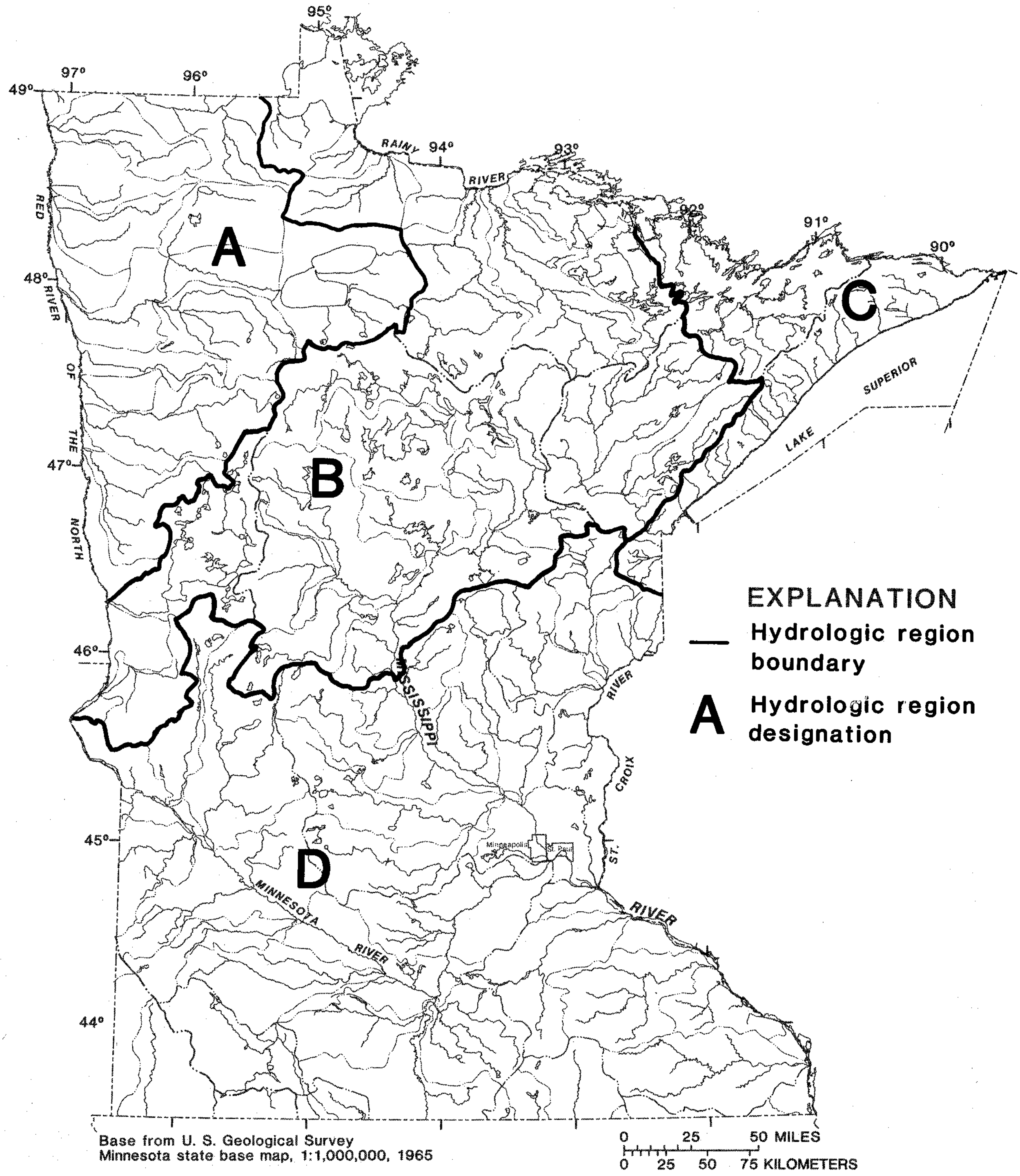


Figure 3.--Hydrologic regions in Minnesota.

Basin Characteristics Investigated

Necessary preconditions for development of useful transfer equations are that the basin characteristics used be limited in number and easily determined from maps. The basin characteristics (independent variables) investigated for this analysis were contributing drainage area, main channel slope, percent area of storage, percent area of lakes and of forest cover, basin shape, mean annual runoff, and 30-year-normal annual precipitation (Baker and Kuehnast, 1978). Regression analysis indicated that both runoff and precipitation were significant characteristics in some regions. However, runoff generally produced a more linear fit than did precipitation. For that reason, mean annual runoff was used rather than normal annual precipitation in the regression equations.

Definitions and procedures for calculating selected basin characteristics are given in the glossary. Figure 4 shows the mean annual runoff for the State of Minnesota.

ESTIMATING FLOOD FREQUENCY OF UNREGULATED STREAMS

The most reliable estimates of flood-flow magnitudes for specific recurrence intervals are based on an analysis of recorded floods at the site under consideration. Such records are not available at most places of interest, and estimates of floods must be obtained by transfer of information from gaged sites or by analysis of generalized flood-frequency relations. The following is a discussion of both techniques along with example computations.

Analyses of Ungaged Sites using Generalized Relationships

Equations, obtained from multiple-regression analyses of gaging-station data in each hydrologic region, can be used to obtain flood-frequency estimates for ungaged sites on unregulated streams. Peak discharges for selected recurrence intervals can be computed from the empirical equations that relate flood magnitude to basin characteristics. A set of equations to estimate flood peaks for 2-, 5-, 10-, 25-, 50-, and 100-year recurrence intervals (identified as Q_2, Q_5, Q_{10}, etc.) are provided in table 1 for each of the four hydrologic regions in Minnesota. The four regions of the State are outlined on figure 3.

The regions represent areas in which each of the stations in the region exhibits an unbiased response to the equations for that region. However, regional boundaries cannot be precisely defined and particular care should be exercised when the site in question has basin characteristics that differ from the general characteristics of a region. The limiting basin characteristics for each region are discussed in the section on Accuracy and Limitations of Estimating Techniques. If an estimate is to be made downstream from a regional boundary that a stream crosses (the St. Louis River is the only such stream in the State), the discharge at the site should be determined by transfer from the gaged site.

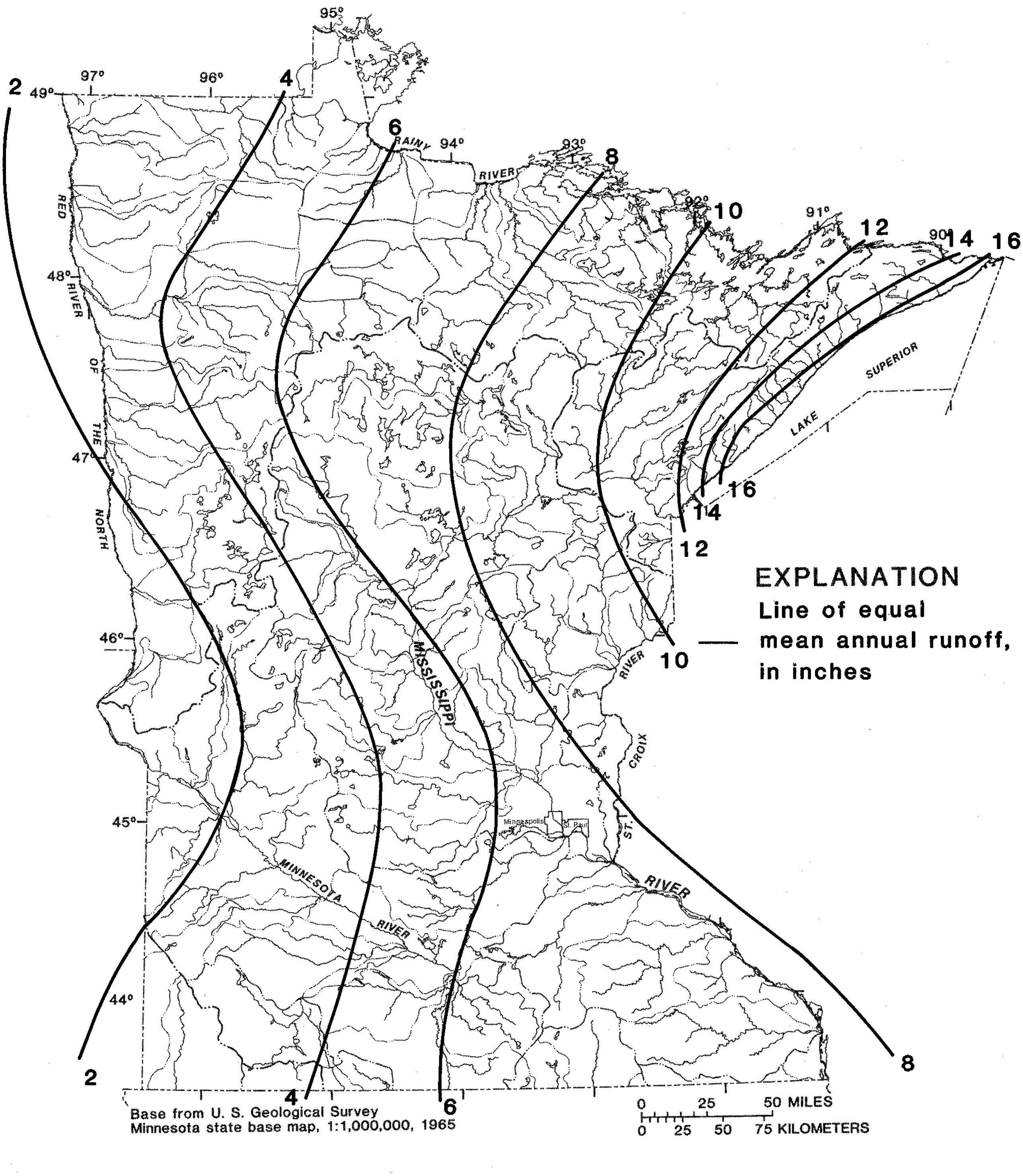


Figure 4.--Mean annual runoff in Minnesota.
(Modified from Baker, Nelson, and Kuenast, 1979, p. 6)

Table 1.--*Prediction equations, standard errors of the estimate (SEE), and equivalent years of record (EY) for all regions*

Prediction equations	SEE (percent)	EY
Region A (39 stations used)		
$Q2 = 28.2\ A^{0.616}\ (St+1)^{-0.108}$	36	5.5
$Q5 = 62.3\ A^{0.617}\ (St+1)^{-0.186}$	37	6.1
$Q10 = 92.5\ A^{0.615}\ (St+1)^{-0.227}$	40	6.7
$Q25 = 139.\ A^{0.613}\ (St+1)^{-0.270}$	45	7.5
$Q50 = 179.\ A^{0.610}\ (St+1)^{-0.298}$	49	7.5
$Q100 = 224.\ A^{0.608}\ (St+1)^{-0.323}$	53	7.5
Region B (41 stations used)		
$Q2 = 2.98\ A^{0.843}\ (Lk+1)^{-0.531}\ R^{0.902}$	33	3.8
$Q5 = 8.88\ A^{0.836}\ (Lk+1)^{-0.587}\ R^{0.654}$	39	3.4
$Q10 = 14.8\ A^{0.833}\ (Lk+1)^{-0.612}\ R^{0.544}$	43	3.6
$Q25 = 24.5\ A^{0.829}\ (Lk+1)^{-0.636}\ R^{0.444}$	48	4.2
$Q50 = 33.1\ A^{0.827}\ (Lk+1)^{-0.651}\ R^{0.387}$	51	4.3
$Q100 = 42.7\ A^{0.825}\ (Lk+1)^{-0.662}\ R^{0.342}$	54	4.5

Table 1.--*Prediction equations, standard errors of the estimate (SEE), and equivalent years of record (EY) for all regions*--continued

Prediction equation	SEE (percent)	EY
Region C (27 stations used)		
$Q2 = 20.3\ A^{0.856}\ (St+1)^{-0.327}\ S^{0.288}$	49	1.4
$Q5 = 24.1\ A^{0.851}\ (St+1)^{-0.339}\ S^{0.383}$	50	1.9
$Q10 = 24.3\ A^{0.852}\ (St+1)^{-0.338}\ S^{0.451}$	50	2.5
$Q25 = 23.0\ A^{0.855}\ (St+1)^{-0.333}\ S^{0.536}$	51	3.4
$Q50 = 21.4\ A^{0.858}\ (St+1)^{-0.326}\ S^{0.599}$	51	4.1
$Q100 = 19.7\ A^{0.862}\ (St+1)^{-0.318}\ S^{0.660}$	52	4.7
Region D (139 stations used)		
$Q2 = 3.24\ A^{0.738}\ (St+1)^{-0.377}\ S^{0.302}\ R^{1.08}$	43	4.5
$Q5 = 7.92\ A^{0.732}\ (St+1)^{-0.392}\ S^{0.324}\ R^{0.937}$	44	5.3
$Q10 = 12.3\ A^{0.728}\ (St+1)^{-0.401}\ S^{0.335}\ R^{0.869}$	47	6.1
$Q25 = 19.5\ A^{0.723}\ (St+1)^{-0.409}\ S^{0.347}\ R^{0.801}$	52	7.1
$Q50 = 25.9\ A^{0.720}\ (St+1)^{-\ .415}\ S^{0.355}\ R^{0.760}$	56	7.2
$Q100 = 33.1\ A^{0.716}\ (St+1)^{-0.419}\ S^{0.362}\ R^{0.724}$	60	7.3

The basin characteristics used in the estimating equations include drainage area (A), slope (S), percent storage (St), percent lakes (Lk), and mean annual runoff (R). Values for the basin characteristics should be determined by the methods described in the glossary.

Various combinations of basin characteristics define the significant independent variables in the equations for the regions (table 1). Years of record is used to determine a weighting factor for calculating peak discharges at sites on gaged streams. The standard error of the estimate and equivalent years of record are discussed in the section "Accuracy and Limitations of Estimating Techniques".

The use of regression equations to estimate flood discharges on ungaged streams is explained in the example below. The technique is similar for all regions and all recurrence intervals for which equations are provided. If an estimated discharge is required for a recurrence interval not defined by an equation, plots of frequency-curve values for the site can be obtained by solution of equations for all intervals. The desired discharge then may be estimated graphically.

Example 1

Estimate the 25-year peak discharge for an ungaged site on Spring Creek in Swift County, at the crossing of State Highway 9, 3 1/2 miles west of Sunburg.

1. Inspection of figure 3 shows that the site is located in Region D.

2. Inspection of table 5 indicates that no gaging-station data are available for this stream; therefore, flow-frequency estimates should be derived from regional equations. The appropriate equation for the 25-year flood is found in table 1.

3. Drainage area above the point of interest is outlined on the De Graff SE 7-1/2-minute topographic map. The drainage area (A) is planimetered as 1.28 mi^2.

4. Total lake, pond, and swamp area is determined from the map by the grid method described in the discussion on storage in the glossary. Fifteen of the small grid squares are counted as storage area; the area computation follows:

   $$15 \text{ squares} \times 0.00144 \text{ mi}^2 = 0.02 \text{ mi}^2$$

   Percent storage is computed by dividing the storage area by the drainage area and multiplying by 100:

   $$\text{Storage} = 0.02/1.28 \times 100 = 1.6 \text{ percent}$$

5. The mean annual runoff, estimated from figure 4 is 3.0 inches.

6. The main channel slope is computed as follows:

The length of the main channel is measured to be 1.49 miles to the end of the watershed divide.

Elevation at mile 0.15 (10% of channel length) is 1,212 ft

Elevation at mile 1.27 (85% of channel length) is 1,235 ft

Main channel slope is (1235-1212)/1.12 = 20.5 ft/mi.

7. Region D equation for 25-year flood (table 1):

$$Q_{25} = 19.5A^{0.723}(St+1)^{-0.409}S^{0.347}R^{0.801}$$

$$Q_{25} = 19.5(1.28)^{0.723}(2.6)^{-0.409}(20.5)^{0.347}(3.0)^{0.801}$$

$$Q_{25} = 19.5x1.195x.6765x2.852x2.411$$

$$Q_{25} = 108 \text{ ft}^3/\text{s} \ (3.06 \text{ m}^3/\text{s})$$

Although the discharge is reported to three significant digits, the accuracy of the estimating technique does NOT warrant confidence to this degree of accuracy. The 95-percent confidence limits are 40 ft^3/s to 290 ft^3/s.

Transfer of Flood-Flow Information from a Gaged Site

The flood characteristics defined by frequency analyses of gaging-station records listed in tables 2 through 5 (tables 2 - 5 are at end of the report) provide estimates for ungaged locations near a station, particularly where long-term records are available. Transfer of defined flow-frequency information to upstream or downstream sites on the same stream should be accomplished by an adjustment factor that is a function of the basin characteristics showing significant intrabasin variation. The transfer equation may be a function of several characteristics; however, the equation always includes drainage area. Consequently, frequency data are transferred most often using the following equation:

$$Q_{T,u} = Q_{T,g}(A_u/A_g)^B$$

where:

$Q_{T,u}$ is the T-year flood-magnitude estimate for the ungaged site,

$Q_{T,g}$ is the computed T-year flood-frequency value for the gaged site (from tables 2 to 8),

A_u is the drainage area for the ungaged site,

A_g is the drainage area for the gaged site (from tables 2 to 8), and

B is the exponent for drainage area for the T-year flood obtained from the regression equation for the region in which the site is located.

Use of the transfer relation should be limited to sites that differ in area by no more than 50 percent from the area of the gaged site. If other characteristics of the ungaged site are significantly different from the gaged site, the T-year flood should be multiplied by the ratio of the basin characteristics of ungaged and gaged sites raised to the regional exponent of that characteristic for the T-year flood (see example 3). Where the period of record is shorter than 20 years, flood-frequency estimates should be based on transfer of information from gaged sites and results from the regional estimating equations.

Example 2

Estimate the 50-year flood on the Sauk River at Cold Spring, an ungaged site.

1. Inspection of figure 2 and table 5 indicate the availability of gaging-station data for the Sauk River in close proximity to Cold Spring. The station is identified as Sauk River near St. Cloud, (station no. 05270500). Because the gaged site has 55 years of record, a simple transfer of peak flow will provide an accurate estimate for the ungaged site.

2. A contributing drainage area of 832 mi^2 at Cold Spring is obtained by planimetering topographic maps.

3. The reduction in drainage area at Cold Spring is only 10 percent from the 925 mi^2 listed in table 5 for the St. Cloud gaging station. Therefore, a transfer of flood characteristics by drainage area ratio is appropriate. Visual inspection of the topographic maps indicates that the other significant basin characteristics are similar to the gaged site. Area only will be used to determine the 50-year-peak flood at Cold Spring.

4. Region B exponent for drainage area for the 50-year flood is 0.827.

5. From table 4, $Q_{50,g} = 6,300\ ft^3/s$.

6. By substitution into the transfer equation:

$$Q_{50,u} = Q_{50,g}(A_u/A_g)^{(0.827)}$$

$$Q_{50,u} = 6,300(832/925)^{0.827}$$

$$Q_{50,u} = 6,300 x 0.916$$

$$Q_{50,u} = 5,770\ ft^3/s\ (153\ m^3/s)$$

Example 3

Estimate the 10-year flood on Silver Creek at the crossing of County Highway 11 east of Rochester.

1. Inspection of figure 2 and table 5 reveals a gage on Silver Creek, station number 05372950, downstream from the selected site. Because the gaged site has only 15 years of record, a weighted estimate of the flood discharge should be made. The proper equation for this site, located in region D, is found in table 1.

2. The drainage area, outlined on the Chester, Minnesota, 7-1/2-minute topographic map, is planimetered and found to be 9.87 mi^2.

3. The main channel is extended to the drainage boundary and the length is measured at 6.28 mi. The 10- and 85-percent points are located on the map and the elevations interpolated. The difference in elevation is 1290-1150 = 140 ft and contributing main channel length is 0.75x6.28 = 4.71 mi, thus:

$$S = 140/4.71 = 29.7 \text{ ft/mi.}$$

4. Total storage is determined by counting the grid squares of pond area within the drainage boundary and multiplying by the area per square.

$$St = 2 \text{ squares x } 0.00144 = 0.003 \text{ mi}^2$$

$$St = (0.003/9.87)\text{x}100 = 0.03 \text{ percent.}$$

5. Mean annual runoff is estimated to be 7.2 inches from figure 4.

6. The 10-year flood is estimated using the regression equations for Region D:

$$Q_{10} = 12.3A^{0.728}(St+1)^{-0.401}S^{0.335}R^{0.869}$$

$$Q_{10} = 12.3(9.87)^{0.728}(1.03)^{-0.401}(29.7)^{0.335}(7.2)^{0.869}$$

$$Q_{10} = 1{,}114 \text{ ft}^3/\text{s } (31.5 \text{ m}^3/\text{s})$$

7. The 10-year flood is estimated by transfer of data from the gaged site, using the basin characteristics of slope and drainage area, which change significantly between sites. Thus, the computed 10-year flood at the gaged site will be factored by the ratio of the areas raised to the 0.728 power times the ratio of slopes raised to the 0.335 power.

$$Q_{10,u} = Q_{10,g}\ (A_u/A_g)^{0.728}\ (S_u/S_g)^{0.335}$$

$$Q_{10,u} = 1{,}990\ (9.87/17.3)^{.728}\ (29.7/32.3)^{.335}$$

$$Q_{10,u} = 1{,}285 \text{ ft}^3/\text{s } (36.4 \text{ m}^3/\text{s})$$

8. The discharges computed in steps 6 and 7 above are weighted and combined to yield an adjusted 10-year flood estimate. The weighting factors used are the equivalent years of record for the regression estimate (step 6) and the actual years of record for the estimate from the gaged site (step 7). The equivalent years of record for estimating a 10-year flood by regression equation in region D is 6.1 from table 1. The number of years of record at the gaged site is 15 from table 5.

$$Q_{10} = (6.1x1,114+15x1,285)/(6.1+15)$$

$$Q_{10} = 14500/21.1 = 1,240 \text{ ft}^3/\text{s} \ (35.1 \text{ m}^3/\text{s})$$

ACCURACY AND LIMITATIONS OF ESTIMATING TECHNIQUE

The accuracy of a statistically defined equation is measured by the closeness of the estimated value to the true value. The U.S. Water Resources Council (1981a, p. 48-49) describes two elements of accuracy, variance, and bias. Variance is a measure of the random variation about the mean of the estimate, and bias is the deviation of the mean of the estimate from the true value of the mean.

Random variation about the mean is caused by a combination of factors. Three of the most significant factors are discussed below. Errors in predicting flood-flow magnitudes result from short sampling records, which may not be a representative sample of the population of annual peaks, and from the assumptions made in procedures for defining the magnitude of flood flows. Errors also result from the inability to completely describe drainage-basin characteristics. No matter how complete the description of a drainage basin, differences exist that contribute, in varying degrees, to the runoff characteristics of a basin. As an example, morphologic features such as storage may be described as a statistic (percent storage), but the impact of each area of storage, its size or relative position in the drainage basin, cannot be accounted for completely. The third source of random error is a result of the empirical nature of the model. The assumptions of a linear-regression model may not be adequately met even though every effort is made to reduce departures from the assumptions.

Bias of an estimate may result from bias in the dependent variable or from an inadequate statistical model. Any bias in the dependent variable (the T-year flood discharge) is most likely the result of time-sampling error. Because most of the data used in this analysis are based on gages operated between 1958 and 1983, the derived flood statistics reflect that period of time and may or may not be a representative sample of the entire population. The statistical model also may be biased because the assumptions of linear regression are not adequately met or because of misspecification of the independent variables. The equations obtained by linear-regression techniques may contain extraneous independent variables or a significant variable may have been omitted.

The accuracy of an estimate made using a statistically derived equation is a function of the accuracy described above and also random variations produced by different users of that equation. Each user or planner will make certain decisions based on his or her best judgment about the actual outline of the drainage basin, what constitutes storage, tracing the path of the main channel, and interpolating values from contour lines. These decisions introduce additional random variations into the model that are not accounted for by the statistical techniques used. The variations introduced in this way differ from user to user and basin to basin. For example, the drainage-basin outline generally is much easier to define in hilly terrain than in flat terrain because the more closely spaced contours reduce ambiguity in the location of the basin divide. Therefore, the random errors introduced by the hydrologist generally will be smaller where the topography is hilly because the judgments regarding drawing of the drainage outline will be easier and will be more consistent between users than in areas of flat topography.

In general, estimates of the probability of future flood occurrences become more accurate with greater length of record (Hardison, 1969). The standard error of the estimate of a predicted flood magnitude for a given recurrence interval decreases approximately proportionally to the square root of the length of gaging station record. At or near gage sites, flood characteristics may be based on analysis of actual records collected at the gage (from tables 2 to 5) or may be computed from regional estimating relations. Weighted averages of flood estimates by regression equations and by transfer relations generally are used when the percent change in any basin characteristic exceeds 10 percent or when the period of gaging is less than 20 years.

The standard error of the estimate is a measure of the distribution of the observed data about the regression surface. The standard error, reported in percent of estimated flow, is the range of deviations from the regression surface to be expected approximately two-thirds of the time. Because the variables used in these analyses are expressed in logarithmic form, the percent standard errors are larger in the positive direction. The values of percent standard error reported in table 1 are the average values. The equivalent number of years of record is an estimate of the information obtained from the regression equation when applied to an ungaged site. In other words, an estimate of a flood based on a regression equation is approximately as good as that obtained from a gaged site operated for that period of time. Hardison (1971) presents an equation that defines the equivalent years of record represented by a regression equation. The equivalent number of years of record is computed from the ratio of the mean variance of the logs of the annual peaks to the mean square error of the regression, multiplied by a factor dependent on the return period and mean coefficient of skewness.

Flood-frequency relations expressed in this report may be used to estimate magnitude and frequency of floods on most Minnesota streams. The applicability and accuracy of these relationships depends on whether the basin characteristics above the site under consideration are within the range of characteristics used to define the frequency relations. The range in sampled basin characteristics is large enough to allow use of the frequency relations at most sites where streamflow is not significantly affected by regulation, diversion, or urbanization. The acceptable range for each of the physical characteristics to be considered is tabulated in table 6 (table 6 is at end of the report). Where runoff is included as an independent variable in an equation, the sampling is complete enough to ensure that the entire range of values may be used.

Corrections must be made at sites immediately below a lake or ponding area where the storage capacity is large in relation to total drainage area and could seriously alter flood characteristics. In such places, the frequency relations may be used as an aid in developing a hydrograph of inflow for use in routing flow through the storage area to the site.

SUMMARY AND CONCLUSIONS

This report contains an analysis of significant flood information for Minnesota streams, except for miscellaneous flood measurements, and very short or unpublished records (see tables 2 to 5). Flood-frequency analyses of the annual series peak data from 246 gaging stations were used to investigate regional relations. A regression analysis of the regionalized data relates peak flows to basin characteristics. The resulting regional equations can be used to estimate flood flows at ungaged, unregulated sites for 2-, 5-, 10-, 25-, 50-, and 100-year recurrence intervals. Analyses of the standard errors of estimate, regression coefficients, and residuals show that these equations provide good estimates of selected frequency annual series peak flows subject to certain limitations.

The use of weighting factors for stations in the regression analysis provides an expedient technique for reduction of standard error. This technique increases the weight of data from longer-term stations and stations with more accurately defined flood characteristics so that shorter-term records may be incorporated into the analysis without adversely affecting the results.

REFERENCES CITED

Baker, D. G. and Kuehnast, E., L., 1978, Climate of Minnesota (part X - precipitation normals for Minnesota: 1941-1970): Agricultural Experiment Station, University of Minnesota, 16 p.

Baker, D. G., Nelson, W. W., and Kuehnast, E. L., 1979, Climate of Minnesota (part XII - The Hydrologic Cycle and Soil Water, Technical Bulletin 322): Agricultural Experiment Station, University of Minnesota, 23 p.

Benson, M. A., 1962, Factors influencing the occurrence of floods in a humid region of diverse terrain: U.S. Geological Survey Water-Supply Paper 1580-B.

Bevington, P. R., 1969, Data reduction and error analysis for the physical sciences: McGraw-Hill, Inc., New York, N. Y., 336 p.

Bobee, B., 1973, Sample error of T-year events computed by fitting a Pearson Type 3 distribution: Water Resources Research 9, p 1264-1270.

Dixon, W. J. and others, 1985, BMDP statistical software manual, 1985 reprinting: University of California Press, Berkely, Calif., 734 p.

Draper, N. R., and Smith, H., 1981, Applied regression analysis: Second Edition, John Wiley and Sons, New York, N. Y., 709 p.

Guetzkow, L. C., 1977, Techniques for estimating magnitude and frequency of floods in Minnesota: U.S. Geological Survey Water-Resources Investigations 77-31, 33 p.

Hardison, C. H., 1969, Accuracy of streamflow characteristics in Geological Survey Research: U.S. Geological Survey Professional Paper 650-D, p D210-214.

____ 1971, Prediction error of regression estimates of streamflow characteristics at ungaged sites: U.S. Geological Survey Professional Paper 750-C, p C228-C236.

Hocking, R. R., 1976, The analysis and selection of variables in linear regression: Biometrics 32, p 1-49.

Linsley, R. K., Kohler, M. A., and Paulhus, J. L. H., 1982, Hydrology for Engineers: McGraw-Hill Inc. New York, N. Y., 508 p.

Patterson, J. L., and Gamble, G. R., 1968, Magnitude and frequency of floods in the United States, Part 5: U.S. Geological Survey Water-Supply Paper 1678, 546 p.

Prior, C. H., 1949, Magnitude and frequency of floods in Minnesota: Minnesota Dept. of Conservation, Division of Waters Bulletin 1, 128 p.

Prior, C. H., and Hess, J. H., 1961, Floods in Minnesota, Magnitude and frequency: Minnesota Dept. of Conservation, Division of Waters Bulletin 12, 142 p.

U.S. Water Resources Council, 1981a, Estimating peak flow frequencies for natural ungaged watersheds, a proposed nationwide test: Hydrology Committee, 346 p.

____ 1981b, Guidelines for determining flood flow frequency, Revised September, 1981: Hydrology Committee Bulletin 17B, 231 p.

Wiitala, S. W., 1965, Magnitude and frequency of floods in the United States, Part 4: U.S. Geological Survey Water-Supply Paper 1677, 357 p.

GLOSSARY

Basin shape (Sh) - Conceptually, the ratio between width and length of the basin, dimensionless. To compute the value for basin shape the area of basin is divided by the main channel length squared.

Drainage area (A) - The area contributing directly to runoff, in square miles. An outline of the drainage area should be drawn on topographic maps and the outline planimetered to determine the area. When available, 7 1/2-minute or 15-minute quadrangle maps should be used.

Forest cover (F) - Forest cover is defined on topographic maps, expressed as a percentage of the contributing drainage area. Forest cover can most easily be determined with a transparent grid. The grid is placed on the map and the number of squares covering forested areas are added together and multiplied by the area of each square, calculated at the scale of the map being used. (A transparent grid suitable to most map scales is enclosed in a packet at the back of this report.)

Lake storage (Lk) - Lake storage can be computed in a manner similar to that used to determine forest cover. Expressed as a percentage of the contributing drainage area.

Main channel slope (S) - Mean slope of the channel computed between points 10 and 85 percent of the main channel length upstream from the point of interest, in feet per mile. The main channel length is defined as the stream bed extending from the site to the basin divide. The 10- and 85-percent points are located on the map and the elevations of the points are interpolated from the topographic contours.

Mean annual runoff (R) - Average annual runoff during 1960-76, in inches. From Climate of Minnesota, Part XII-The Hydrologic Cycle and Soil Water, published by the Agricultural Experiment Station, University of Minnesota. The value for this basin characteristic is determined by locating the centroid of the basin on the map and interpolating between isolines. A copy of this map is found on figure 4.

Storage (St) - The storage area includes all lakes, ponds, and wetlands in the basin, expressed as a percentage of the contributing drainage area. The easiest technique to determine storage is by use of the grid method described above for forest cover.