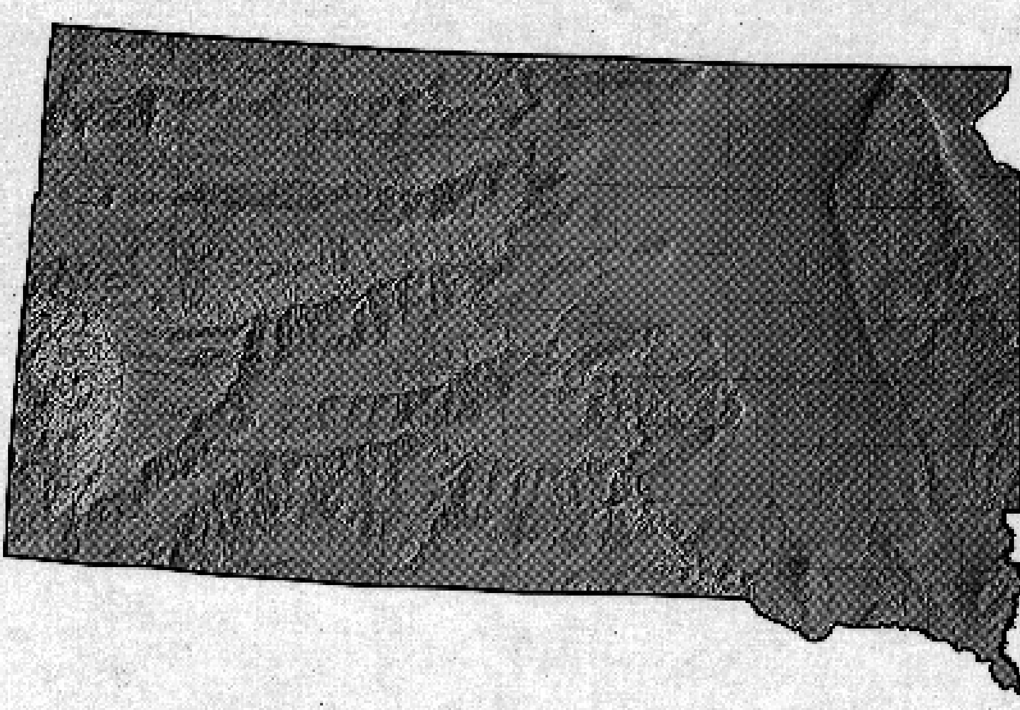
Prepared in cooperation with the
South Dakota Department of Transportation

Techniques for Estimating Peak-Flow Magnitude and Frequency Relations for South Dakota Streams

Water-Resources Investigations Report 98-4055



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

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

Techniques for Estimating Peak-Flow Magnitude and Frequency Relations for South Dakota Streams

By Steven K. Sando

605-353-7176 x230

Water-Resources Investigations Report 98-4055

Prepared in cooperation with the South Dakota Department of Transportation

U.S. DEPARTMENT OF THE INTERIOR
BRUCE BABBITT, Secretary

U.S. GEOLOGICAL SURVEY
Thomas J. Casadevall, Acting Director

United States Government Printing Office: 1998

For addtional information write to:

District Chief
U.S. Geological Survey
1608 Mt. View Rd.
Rapid City, SD 57702

Copies of this report can be purchased from:

U.S. Geological Survey
Branch of Information Services
Box 25286
Denver, CO 80225-0046

CONTENTS

PLATE

ILLUSTRATIONS

TABLES

Techniques for Estimating Peak-Flow Magnitude and Frequency Relations for South Dakota Streams

By Steven K. Sando

ABSTRACT

A generalized skew coefficient analysis was completed for South Dakota to test the validity of using the generalized skew coefficient map in Bulletin 17B of the 1982 United States Water Resources Council, "Guidelines for Determining Flood Flow Frequency." Results of the analysis indicate that the Bulletin 17B generalized skew coefficient map generally provides adequate generalized skew coefficients for estimating peak-flow magnitudes and frequencies for South Dakota gaging stations.

Peak-flow records through 1994 for 197 continuous- and partial-record streamflow-gaging stations that had 10 or more years of unregulated systematic record were used in a generalized least-squares regression analysis that relates peak flows for selected recurrence intervals to selected basin characteristics. Peak-flow equations were developed for recurrence intervals of 2, 5, 10, 25, 50, 100, and 500 years for seven hydrologic subregions in South Dakota. The peak-flow equations are applicable to natural-flow streams that have drainage areas less than or equal to 1,000 square miles. The standard error of estimate for the seven hydrologic subregions ranges from 22 to 110 percent for the 100-year peak-flow equations.

Weighted peak flows for various frequencies based on gaging-station data and the regional regression equations are provided for each gaging station. Examples are given for (1) determining peak-flow magnitudes and frequencies for ungaged sites on ungaged streams; (2) determining weighted peak-flow magnitudes and frequencies for gaging stations; and (3) using the drainage-area ratio method for determining peak-flow magnitudes and frequencies for ungaged sites near a gaging station on the same stream and ungaged sites between two gaging stations on the same stream.

INTRODUCTION

The magnitude and frequency of peak flows are essential parameters for designing engineering structures (such as bridges, levees, and culverts), land-use planning, establishing rates for flood insurance, and developing emergency evacuation plans for flood-prone areas. Accurate estimates of peak-flow magnitude for various frequencies (recurrence intervals) are necessary for effective structural design and planning purposes. Underestimates of peak-flow magnitudes may result in disruption of service, costly maintenance, and loss of life, while overestimates may result in excessive construction costs. Design and planning activities often require peak-flow magnitude and frequency information at locations where no or inadequate peak-flow data have been collected. Therefore, methods are needed to provide accurate estimates of peak-flow magnitude for various recurrence intervals, such as 2, 5, 10, 25, 50, 100, and 500 years, at ungaged sites. The U.S. Geological Survey (USGS) conducted a study in cooperation with the South Dakota Department of Transportation to develop techniques for estimating peak-flow magnitudes and frequencies at ungaged sites.

Peak-flow data collected at continuous-record and crest-stage gages over a period of years are used to estimate peak-flow magnitudes and frequencies at gaged sites. Peak-flow frequencies at gaged sites commonly are determined by fitting a probability distribution function to the series of annual peak flows. All Federal agencies and many State agencies and local consultants use the log-Pearson Type III distribution when determining peak-flow magnitude and frequency relations for gaging stations and follow the procedures described in Bulletin 17B, "Guidelines for Determining Flood Flow Frequency" (United States Water Resources Council, 1982; hereinafter referred to as Bulletin 17B). The procedures described in Bulletin 17B were used to determine peak-flow frequencies in this report.

The network of continuous-record and crest-stage gages operated by the USGS in South Dakota provide regional peak-flow information that can be used to estimate peak-flow relations at ungaged sites. Peak-flow frequencies at ungaged sites on ungaged streams commonly are estimated from equations developed using regression procedures that relate peak-flow magnitudes for various frequencies at gaging stations to basin and climatic characteristics. Peak-flow frequencies at ungaged sites on gaged streams can be estimated using logarithm-proportional drainage-area relations applied to nearby gaged sites.

Purpose and Scope

The purposes of this report are (1) to present the results of an analysis of the validity of using the generalized skew coefficient map from Bulletin 17B for South Dakota; (2) to publish peak-flow magnitudes for selected recurrence intervals from 2 to 500 years for gaging stations with 10 or more years of generally unregulated systematic record; (3) to present regression equations for estimating peak-flow magnitude and frequency relations for ungaged sites on ungaged unregulated streams that have drainage areas of 1,000 mi^2 (square miles) or less; (4) to present a procedure for computing peak-flow magnitude and frequency relations for ungaged sites near gaging stations on gaged streams; and (5) to discuss data needs for improving future regional peak-flow frequency analyses in South Dakota.

The generalized skew coefficient and regression analyses were based on data for 197 continuous- and partial-record streamflow gaging stations that had 10 or more years of generally unregulated systematic record. Of these stations, 156 are in South Dakota, 6 are in Iowa, 15 are in Minnesota, 6 are in Montana, 5 are in Nebraska, and 9 are in North Dakota. The locations of the 197 gaging stations are shown in figure 1. Fitting of the log-Pearson Type III distribution using the procedure described in Bulletin 17B was used to compute peak-flow magnitude and frequency relations for the 197 gaging stations. The generalized least-squares (GLS) regression technique (Tasker and Stedinger, 1989) was used to develop equations that can be used to estimate peak-flow magnitude and frequency relations for ungaged sites and to compute weighted peak-flow magnitude and frequency relations for the 197 stations. A drainage-area ratio method was developed to estimate peak-flow magnitude and frequency relations for ungaged sites near gaging stations on gaged streams.

The gaging-station data used in the regression analyses generally reflect unregulated conditions and do not reflect substantial influences from man's activities such as impoundment or urbanization. Peak flows at sites that are substantially influenced by impoundment can be estimated by (1) peak-flow magnitude and frequency analysis on gaging-station data at the site of interest that includes impoundment effects; or (2) estimation of peak flows upstream from the impoundment and application of routing techniques to estimate peak flows at locations downstream from the impoundment. Peak flows at sites that are substantially influenced by urbanization can be estimated using techniques presented in Sauer and others (1983).

Acknowledgments

The author recognizes the hard work and dedication of the many USGS hydrologic technicians that collected the streamflow data on which this report is based. Also, the author appreciates the assistance of the many Federal, State, and local agencies that financially supported the operation of the streamflow gages.

Previous Studies

McCabe and Crosby (1959) used all data available through 1955 and completed a study of the magnitude and frequency of peak flows in North Dakota and South Dakota. Patterson (1966) used data through 1961 and Patterson and Gamble (1968) used data

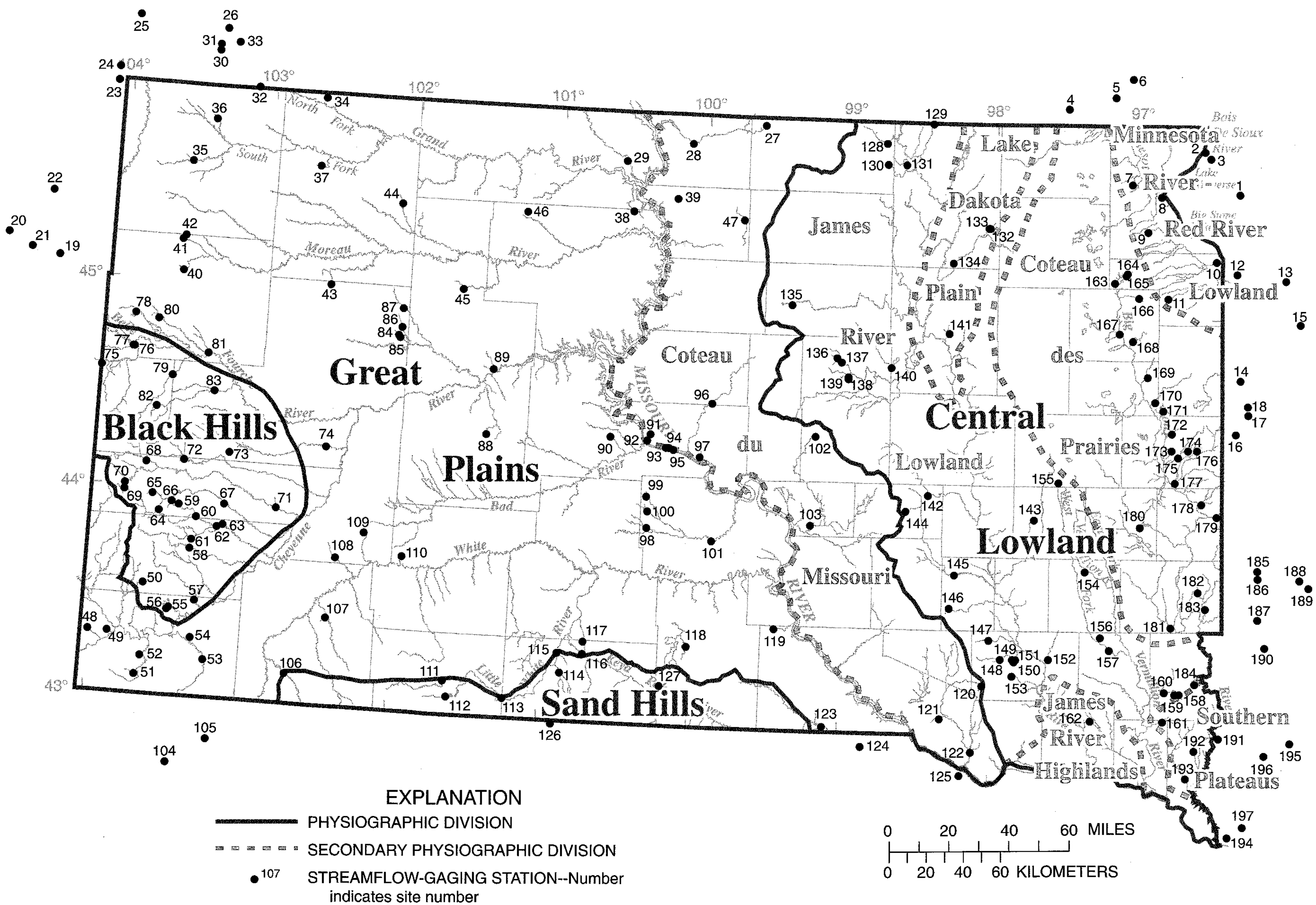


Figure 1. Locations of gaging stations and main physiographic divisions of South Dakota (physiographic divisions modified from Flint, 1955; Fenneman, 1946).

through 1963 to complete parts of a series of reports on the magnitude and frequency of peak flows in the United States. However, the report by Patterson and Gamble (1968) contained little information on peak-flow magnitudes and frequencies for small drainage basins in South Dakota. Larimer (1970) included a limited analysis of magnitude and frequency of peak flows when he evaluated the streamflow-data program for South Dakota, and concluded that a comprehensive study of peak-flow magnitude and frequency in South Dakota was needed. Becker (1974) used data available through 1971 from 162 gaging stations in South Dakota and adjacent states with drainage areas up to 9,000 mi^2 to investigate peak-flow magnitude and frequency relations. Regression equations for estimating peak-flow magnitude for various frequencies at ungaged sites were developed for two regions in South Dakota (eastern region and western region, separated by the western boundary of the James River Basin). Becker (1980) used data available through about 1979 from 123 gaging stations in South Dakota and adjacent states with drainage areas up to about 150 mi^2 to investigate peak-flow magnitude and frequency relations. Becker's 1980 report focused on smaller drainage basins and developed regression equations applicable to the entire State for estimating peak-flow magnitudes for selected frequencies at ungaged sites. Benson and others (1985) presented peak-flow magnitude and frequency for 111 gaged sites based on data available through 1983. Hoffman and others (1986) presented peak-flow magnitude and frequency for 185 gaged sites based on data available through 1985. Burr and Korkow (1996) presented peak-flow magnitude and frequency for 250 gaged sites based on data available through 1994.

General Description of Study Area

The landscape of South Dakota is diverse, with areas of flat prairies, extensive plains, rugged badlands, and the scenic Black Hills. Land-surface elevations range from 7,242 ft (feet) above sea level[1] at Harney Peak in the Black Hills to about 1,000 ft near Big Stone Lake in the northeast corner of the State. A prominent geographic feature of South Dakota is the Missouri River, which bisects the State. The Missouri River drains the entire State, with the exception of a small area in the northeast corner that drains into the Minnesota River or the Red River of the North. There are two major physiographic divisions in South Dakota: the Great Plains and the Central Lowland (fig. 1). The border between the two divisions is the eastern boundary of the Coteau du Missouri. Part of the Sand Hills of Nebraska extends into south-central South Dakota, and the Black Hills are located in southwest South Dakota.

The Great Plains physiographic division is characterized by undulating uplands dissected by stream valleys, sizable flood plains along the larger streams, occasional bluffs and buttes, and areas of badlands (U.S. Geological Survey, 1975). The topography of the Great Plains results from the erosion of easily erodible sedimentary rock. The Great Plains generally is unglaciated, and has well-developed natural drainage systems that form major streams typically flowing west to east. The primary streams draining the Great Plains include the Grand, Moreau, Cheyenne, Bad, and White Rivers. Badlands are most prominent in the White River Basin, but areas typically referred to as "breaks" that have substantial relief commonly occur along or near many of the larger stream channels in western South Dakota. A notable feature within the Great Plains is the Coteau du Missouri, which is the part of the Great Plains that lies east of the Missouri River. It is a hummocky area over which glacial stagnation occurred.

The Central Lowland in eastern South Dakota is an area profoundly influenced by the most recent glaciation. Natural drainage systems in the Central Lowland generally are poorly developed, and numerous lakes and wetlands occur. Typically, major streams in eastern South Dakota flow from north to south. Major streams draining the Central Lowland include the James, Vermillion, Big Sioux, and Little Minnesota Rivers. The Central Lowland primarily consists of the Minnesota River-Red River Lowland, the Coteau des Prairies, the James River Lowland, the Lake Dakota Plain, the James River Highlands, and the Southern Plateaus (Flint, 1955). The low-lying areas of the Minnesota River-Red River Lowland and the James River Lowland are characterized by very flat slopes. The Coteau des Prairies is a massive highland that stands between the Minnesota-Red River Lowland and the James River Lowland. It is drained by the Big Sioux River.

[1]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.

The Sand Hills of Nebraska extend into a small part of south-central South Dakota. Surficial deposits in the Sand Hills region of South Dakota are composed primarily of alluvial terrace and wind-blown sand deposits, all of Quaternary age. The soils in the area have large infiltration capacities atypical of most of South Dakota. The topography of the area primarily consists of ridges and rolling hills formed by erosion of the surface. The Little White and the Keya Paha Rivers are the primary streams draining the Sand Hills region of South Dakota.

The Black Hills are a dome-shaped easternmost uplift of the Wyoming Rocky Mountains. The Black Hills generally consist of metamorphosed and intensely folded sedimentary rocks that rise several hundred feet above the surrounding plains. The topography of the Black Hills is characterized by sharp relief and steep slopes. Stream infiltration to, and resurgent springs discharging from, karstic limestone greatly affect base flow of many of the streams draining the Black Hills. Primary streams draining the Black Hills include Spearfish, Whitewood, Elk, Boxelder, Rapid, Spring, Battle, French, and Beaver Creeks, and the Fall River.

Precipitation and runoff rates in South Dakota differ annually and with season and location. The normal annual precipitation in South Dakota ranges from about 14 in. in the northwest to about 24 in. in the higher elevations of the Black Hills and in the southeast (accessed on the Internet July 3, 1997, at http://www.abs.sdstate.edu/ae/weather/weather.htm). About 70 percent of annual precipitation occurs during the growing season (May through October) and local high-intensity thunderstorms are common. Winter precipitation is small and generally occurs as snow. The average annual runoff ranges from about 0.2 in. in the north-central and extreme southwest to about 2 in. in parts of the Black Hills (Benson, 1986). A large percentage of runoff occurs as a result of snowmelt and rainfall in the spring and early summer.

Peak flows in South Dakota result from both rainfall and snowmelt. No study has been conducted by the USGS to investigate the proportion or relative magnitude of peak flows in South Dakota that occur as a result of snowmelt versus rainfall. A cursory examination of annual peak flows for gaging stations that were used in this regional peak-flow analysis indicated that rainfall-only peaks account for about 65 percent of Central Lowland annual peaks, 85 percent of Great Plains annual peaks, 70 percent of Sand Hills annual peaks, and 90 percent of Black Hills annual peaks. No attempt to separate rainfall-only from snowmelt-influenced annual peak flows was made for this study because of the difficulty in distinguishing between the two types of peaks.

BASIN AND CLIMATIC CHARACTERISTICS

Selected basin and climatic characteristics were used in the analysis of generalized skew coefficients and in the development of regression equations for estimating peak-flow magnitudes for various frequencies for ungaged sites. Basin and climatic characteristics used in these analyses were obtained from the USGS National Water Data Storage and Retrieval System (WATSTORE; Dempster, 1983). The characteristics were chosen based on results from previous regional peak-flow frequency studies and the availability of existing data. Methods for determining basin and climatic characteristics are discussed in more detail in Benson (1962) and Benson and Carter (1973). Many of the values of climatic characteristics included in this study were determined using older publications of what is now the National Weather Service. These data were checked with more recent climatic data to determine whether patterns in climatic variability were well represented by the older data. Although magnitudes of climatic variables sometimes changed substantially when the older publications were compared to recent data, it was determined that the relative variability in the spatial distributions of those variables remained fairly similar. Therefore, the older climatic data were determined to be appropriate for use in this study. The basin and climatic characteristics included in the study are presented in table 1.

ANALYSIS OF GENERALIZED SKEW COEFFICIENT

The Bulletin 17B log-Pearson Type III method requires calculation of a skew coefficient of the logs of the peak flows for a given site. The skew coefficient can be sensitive to extreme events and difficult to accurately estimate at sites with relatively short periods of record. To improve estimates of skew coefficients, Bulletin 17B recommends weighting the station skew with a generalized skew coefficient developed from many long-term stations in an area. The generalized skew coefficient for a site can be estimated using the nationwide map of skew coefficient isolines (lines of equal skew coefficients) presented in Bulletin 17B or by using one of three methods described in Bulletin 17B to estimate generalized skew coefficient with data from at least 20 long-term stations located in the area of interest. For the generalized skew coefficient analysis in South Dakota, 106 stations in and near South Dakota that generally are well distributed and with at least 25 years of record through water year 1994 were used.

The nationwide map of generalized skew coefficients presented in Bulletin 17B was developed in the early 1970's using data from 2,972 gaging stations in the United States with at least 25 years of record. A visual examination of the nationwide skew map indicates that in the northern Great Plains there was a relative paucity of data used to develop the map; only about 20 of the 2,972 stations used to develop the map were in South Dakota. The nationwide mean-squared error of the Bulletin 17B generalized skew coefficient map was reported to be about 0.302. However, using data from the 106 stations used in the South Dakota analysis of generalized skew coefficient, the mean-squared error of the Bulletin 17B nationwide skew map values relative to the station skews was calculated to be 0.484, which is considerably larger than the overall mean-squared error originally calculated for the nationwide skew map.

There is substantial variability in skew coefficients in South Dakota as evidenced by the steep gradient in the skew coefficient isolines for South Dakota on the nationwide map. The generalized skew coefficient isolines generally intersect the State northwest to

Table 1. Selected basin and climatic characteristics used in the South Dakota regional peak-flow frequency analysis

Basin characteristic	Description
Drainage area (A)	Total drainage area, in square miles, including non-contributing areas.
Contributing drainage area (CA)	Drainage area, in square miles, that contributes to surface runoff.
Slope (S)	Main-channel slope, in feet per mile, measured at points 10 and 85 percent of stream length upstream from gage.
Length	Stream length, in miles, measured along channel from gage to basin divide.
Elevation (E)	Mean basin elevation, in feet above mean sea level, measured from topographic maps by transparent grid sampling method.
Percent storage	Area of lakes, ponds, and swamps in percent of contributing drainage area, measured by the grid sampling method.
Percent forest	Forested area, in percent of contributing drainage area, measured by the grid sampling method.
Soil-infiltration index (SII)	Soil-infiltration index, in inches; a relative measure of potential infiltration (soil-water storage), from Natural Resources Conservation Service.
Gaging-station latitude	Latitude of stream-gaging station, in decimal degrees.
Gaging-station longitude	Longitude of stream-gaging station, in decimal degrees.
Mean annual precipitation	Mean annual precipitation, in inches, from U.S. Weather Bureau (1959a).
Precipitation intensity index (PII)	24-hour precipitation intensity, in inches, with a recurrence interval of 2 years (estimated from U.S. Weather Bureau, 1961) minus 1.5.
March 15 snow-cover water equivalent of 2-year recurrence interval	Maximum water equivalent, in inches, of snow cover as of March 15 with a recurrence interval of 2 years (estimated from U.S. Weather Bureau, 1964).
Mean minimum January temperature	Mean minimum January temperature, in degrees Fahrenheit, from U.S. Weather Bureau (1959a).
Mean annual lake evaporation	Mean annual lake evaporation, in inches, from U.S. Weather Bureau (1959b).

southeast and range from +0.6 in the southwest part of the State to -0.4 in the northeast part of the State. The orientation of the isolines generally corresponds to large hydrologic and physiographic differences across South Dakota. For example, the steep forested Black Hills occur in the southwest part of the State. East and north of the Black Hills, the unglaciated Great Plains, with generally well-developed drainage systems, extends to about the Missouri River. The glaciated Missouri Coteau and Central Lowlands with poorly developed natural drainage systems, flat slopes, and numerous lakes and wetlands are east of the Missouri River. The large variability in hydrology and physiography across South Dakota probably contributes to the relatively large variability in skew coefficients. The small number of South Dakota gages used to develop the nationwide skew coefficient map may not have been adequate to define generalized skew coefficient in South Dakota.

Bulletin 17B recommends the use of generalized skew coefficients developed from detailed studies using pooled information from nearby long-term stations instead of generalized skew coefficients taken from the nationwide map. Three methods for estimating generalized skew coefficients are described in Bulletin 17B: (1) plotting station skew coefficients on a map and drawing lines of equal values; (2) developing regression equations that relate station skew coefficients to selected basin and climatic characteristics; and (3) averaging station skew coefficients within a region (or subregions) of interest.

Sampling error, primarily due to differences in record length, introduces a bias into skew coefficient estimates. Tasker and Stedinger (1986) presented a bias correction factor ($1+6/n$; where n = number of years of record) to adjust for the effect of different record lengths on skew estimates. This correction factor was applied to all station skew coefficients prior to the generalized skew analysis.

Station skews for 106 long-term gages used in the South Dakota regional analysis were plotted near the centroid of each respective drainage basin and the resulting map was visually examined for spatial patterns that could be used to draw isolines. Skew coefficients tended to be higher for stations around the Black Hills and in the Little White River Basin, and lower for stations in the eastern part of the State. However, the patterns were not consistent; often there was large variability in skew coefficients between stations in a relatively small area. It was determined that development of isolines would be very difficult and therefore this method was not further pursued.

Regression analysis was performed to determine whether generalized skew coefficients could be accurately estimated using selected basin and climatic characteristics. Initially, station skew coefficients were plotted versus individual basin and climatic characteristics and a correlation analysis was performed to identify potential candidate variables. Eight basin and climatic characteristics (elevation, percent forested area, soil-infiltration index, gaging-station latitude, gaging-station longitude, mean minimum January temperature, maximum water equivalent of snow cover as of March 15 with a recurrence interval of 2 years, and mean annual lake evaporation) were significantly (α=0.05) correlated with the station skew coefficient. All-possible-subsets and stepwise-regression procedures were then performed and the best one-variable up to the best seven-variable models were selected based on Mallow's C_p (Helsel and Hirsch, 1992). From the all-possible-subsets and stepwise-regression results, the most influential explanatory variables were examined for final model selection. Finally, several regressions were performed using the most influential explanatory variables, and the best-fit model was selected based on the PRESS statistic (Helsel and Hirsch, 1992). The best-fit model is

$$G = 1.52\,[\log SII] + 1.04\,[\log E] - 4.52 \qquad (1)$$

where

G = station skew coefficient;
SII = soil-infiltration index; and
E = elevation.

The two explanatory variables in the best-fit model (soil-infiltration index and elevation) showed a weak (r-squared equal 0.20; mean-squared error equal 0.428) but significant relation with the skew coefficient.

The relation between the soil-infiltration index and the skew coefficient was positive and intuitively reasonable. A large soil-infiltration index results in greater infiltration and smaller runoff for many storm events. Thus, basins with a larger soil-infiltration index may have a larger number of annual peak flows concentrated at the lower end of the distribution. However, very large or intense storm events will exceed the soil-infiltration capacity and generate large runoff events. These characteristics could yield a peak-flow distribution with a positive skew.

The relation between elevation and skew coefficient is less obvious. It is possible that elevation serves as a surrogate variable and represents the influence of several physiographic and hydrologic variables. Elevation generally is largest in the southwest part of South Dakota and decreases moving east and north across the State. This distribution generally corresponds to variability in physiography in the State moving from the Black Hills in the southwest, across the unglaciated Great Plains to the glaciated Coteau du Missouri and Central Lowland. The mean-squared error for the best-fit regression equation of 0.428 is relatively high for an estimating equation, but it is slightly better than the mean-squared error of 0.484 of the nationwide map for the 106 stations used in the South Dakota generalized skew analysis.

For the third method of estimating generalized skews, the State was divided into four subregions generally based on the boundaries of the Central Lowland, Great Plains, Sand Hills, and Black Hills shown in figure 1, and station skews for gaging stations within each subregion were averaged. Subregions were selected based on regional boundary definitions from previous studies, reasonably similar hydrology, and an attempt to satisfy the minimum requirement of 20 gaging stations per subregion recommended in Bulletin 17B. This minimum requirement was relaxed for the Black Hills influenced and the Sand Hills influenced subregions because hydrologic considerations warranted defining these two areas as separate subregions.

Comparison of results of different methods for estimating generalized skew are presented in table 2. Examination of the mean-squared errors indicates the regression and subregion-averaging methods provided slightly better fits than the Bulletin 17B nationwide map for some regions, and slightly worse fits for others. None of the generalized skew methods was consistently and clearly better than the others.

To determine how the different skew methods might affect peak-flow estimates, peak-flow magnitudes for selected frequencies were calculated for various groupings of stations used in the regional regression analyses using generalized skews determined from the Bulletin 17B nationwide map and by using the regression and subregion averaging methods (fig. 2). The results of this comparison indicated that differences between the peak-flow estimates resulting from the three different generalized skew methods tended to be small; generally, mean differences between the peak-flow estimates using the three methods were less than 10 percent. The Bulletin 17B nationwide map generally resulted in more positive skew estimates, and therefore, slightly larger peak-flow estimates for larger recurrence intervals. Thus, peak-flow estimates using the generalized skew from the Bulletin 17B nationwide map may, on average, provide more conservative peak-flow estimates for engineering design applications. Based on ease of use, generally small differences in resulting peak-flow estimates, appropriateness for design applications, and promotion of consistency in peak-flow estimates made by different agencies and across state boundaries, the Bulletin 17B national generalized skew map is considered to provide the best generalized skew estimates for sites in South Dakota.

Table 2. Summary statistics for results of methods for estimating generalized skew coefficients in South Dakota

Gaging station group	Number of stations	Mean station skew	Mean skew estimate using Bulletin 17B map	Mean-squared error of Bulletin 17B map value relative to station skew	Mean skew estimate using regression equation	Mean-squared error of regression equation predicted value relative to station skew	Mean skew estimate using subregion averaging method	Mean-squared error of subregion averaging predicted value relative to station skew
All stations	106	-0.283	-0.059	0.484	-0.288	0.419	-0.283	0.403
Central Lowlands	48	-0.422	-0.330	0.290	-0.443	0.311	-0.422	0.289
Great Plains	37	-0.496	0.088	0.271	-0.381	0.326	-0.496	0.293
Sand Hills influenced	7	0.301	0.510	0.132	-0.112	0.121	0.301	0.113
Black Hills influenced	14	0.461	0.194	1.360	0.405	1.178	0.461	1.342

PEAK-FLOW MAGNITUDE AND FREQUENCY RELATIONS FOR GAGING STATIONS

Peak-flow magnitudes and frequencies for 197 gaging stations on streams that have drainage areas less than or equal to 1,000 mi^2 and with periods of unregulated systematic record of at least 10 years were calculated using Bulletin 17B procedures and are listed in table 6 in the Supplemental Information section at the end of this report. The peak-flow magnitudes and frequencies presented in table 6 are identical to those presented in Burr and Korkow (1996) except for the following: (1) Gaging station 06402000, Fall River at Hot Springs, SD, had part of its systematic record influenced by regulation. For this study, the earlier, unregulated part of the systematic record for that station was used, whereas Burr and Korkow (1996) reported peak-flow magnitudes and frequencies for the later, regulated part of the systematic record. (2) Gaging station 06453255, Choteau Creek near Avon, SD, had the 1984 peak annual flow revised in the USGS peak flow data base after publication of Burr and Korkow (1996). For this study, peak flow magnitudes and frequencies for station 06453255 are based on the systematic record that incorporates the revised 1984 peak. (3) Gaging station 06354860, Spring Creek near Herreid, SD, had the 1987 peak annual flow added to the USGS peak flow data base after publication of Burr and Korkow (1996). For this study, peak flow magnitudes and frequencies for station 06354860 are based on the systematic that incorporates the 1987 peak.

PEAK-FLOW MAGNITUDE AND FREQUENCY RELATIONS FOR UNGAGED SITES ON UNGAGED STREAMS

The procedure for determining peak-flow frequencies at an ungaged site depends on whether the site is an ungaged site on an ungaged stream or is located near a gaging station on the same stream. For ungaged sites on ungaged streams, the regional regression equations developed during this study that relate peak flow for selected recurrence intervals to basin and climatic characteristics should be used. For an ungaged site near a gaging station on the same stream, a drainage-area ratio method should be used to estimate the T-year peak flow (where T indicates the selected recurrence interval). The procedure for estimating T-year peak flows for ungaged sites near a gaging station on the same stream is described in the section of this report Peak-Flow Magnitude and Frequency Relations for Ungaged Sites near a Gaging Station on the Same Stream.

Development of Regression Equations

Equations that relate peak-flow values for selected recurrence intervals to basin characteristics were developed using a generalized least-squares (GLS) regression procedure (Tasker and Stedinger, 1989) instead of using the more conventional ordinary least-squares (OLS) regression procedure. Two assumptions of OLS regression that commonly are violated in regional regression analyses are that annual peak flows have constant variance and are independent from site to site. The constant variance assumption typically is violated because the variance is somewhat dependent on the length and timing of the systematic record, which often varies between stations. The independence assumption commonly is violated due to cross correlation between concurrent peak flows for different stations. The GLS regression procedure takes into consideration the time-sampling error in the peak-flow series and the cross correlation between sites, and thus overcomes the violation of assumptions that is inherent when applying OLS regression to regional streamflow studies. The GLS regression procedure also provides better estimates of the predictive accuracy of peak-flow values that are computed by the regression equations and almost unbiased estimates of the variance of the underlying regression model error (Stedinger and Tasker, 1985).

The basic regression model using the GLS procedure can be represented by the following linear equation:

$$\underline{Y} = \underline{x}\,\underline{B} + \underline{u} \qquad (2)$$

where

$\underline{Y}$ = (n x 1) vector of T-year peak-flow events (dependent variable);

$\underline{x}$ = (n x p) matrix of basin and climatic characteristics (explanatory variables);

$\underline{B}$ = (p x 1) vector to be estimated (regression coefficients); and

$\underline{u}$ = (n x 1) random vector (errors).

The best linear unbiased estimator (b) of the parameter vector B for the T-year event (Stedinger and Tasker, 1985) is

$$\underline{b} = (\underline{x}^T\underline{\Lambda}^{-1}\underline{x})^{-1}\underline{x}^T\underline{\Lambda}^{-1}\underline{Y} \qquad (3)$$

where

$\underline{\Lambda}$ = the unknown covariance (weighting) matrix.

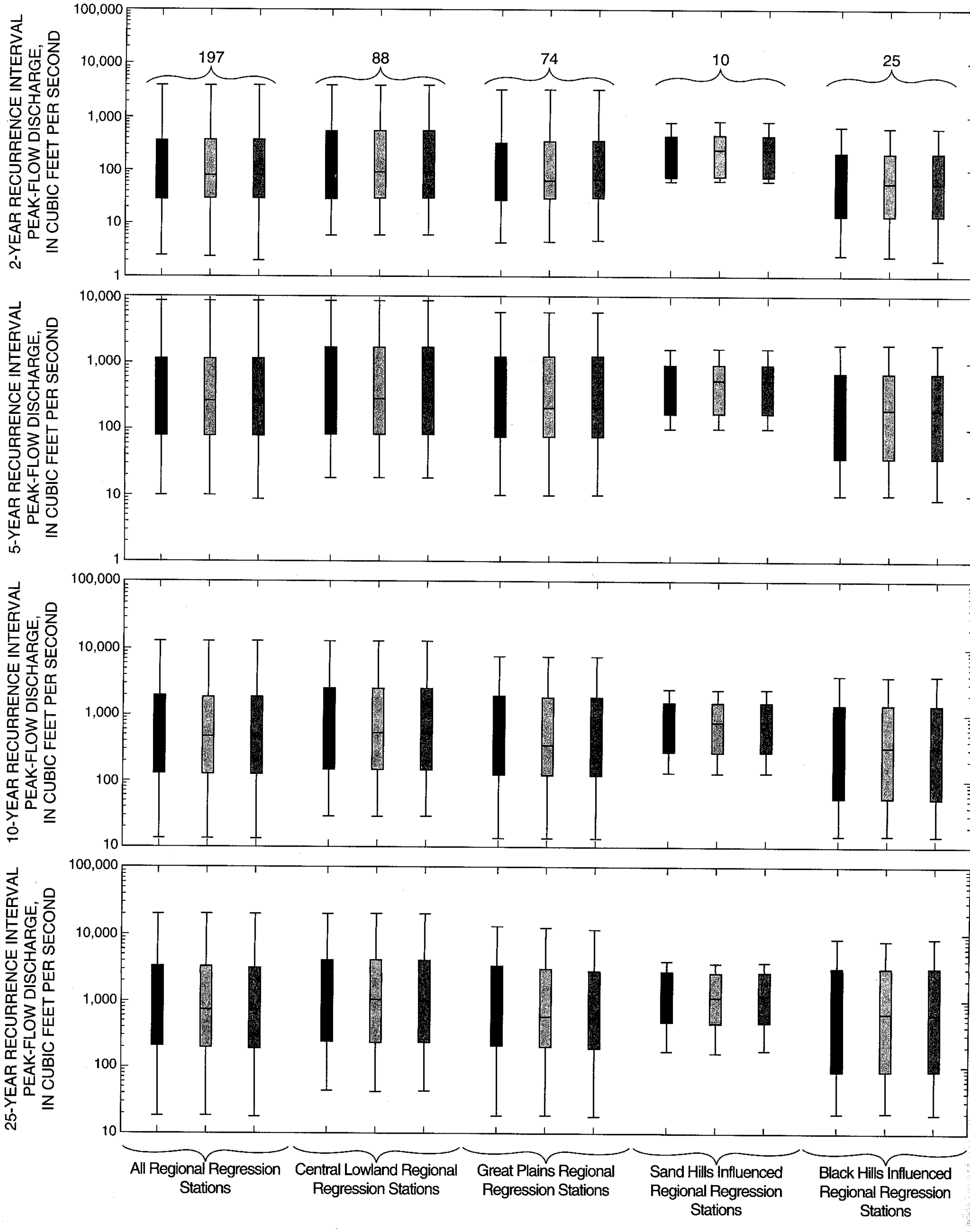


Figure 2. Statistical distributions of peak-flow magnitudes for selected frequencies calculated using three methods for estimating generalized skew coefficient.

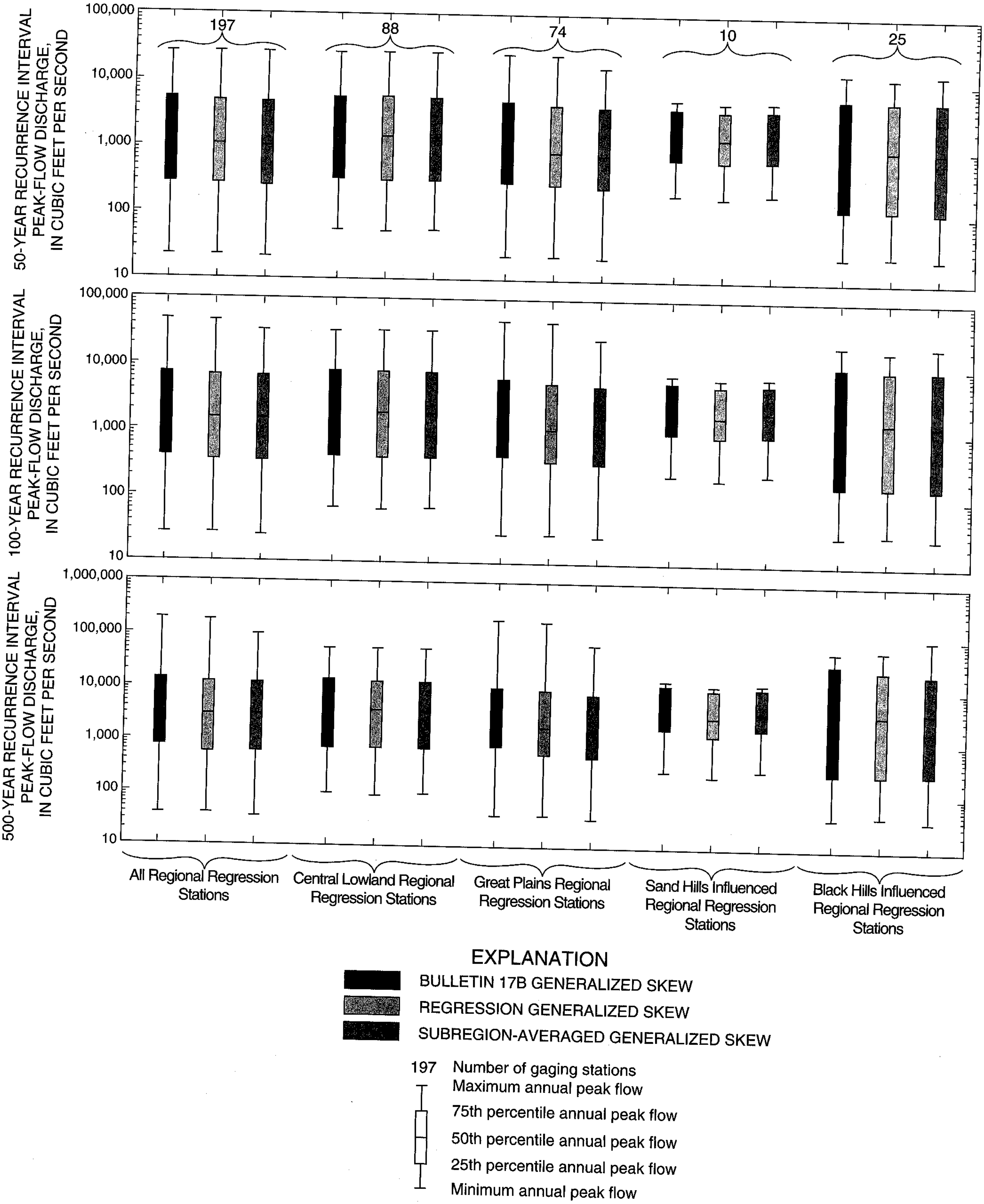


Figure 2. Statistical distributions of peak-flow magnitudes for selected frequencies calculated using three methods for estimating generalized skew coefficient.--Continued

Further details on the GLS procedure and the methods used to determine the unknown matrix are discussed in Stedinger and Tasker (1985) and in Tasker and Stedinger (1989). The computer program GLSNET (G.D. Tasker, written commun., 1997; based on Tasker and Stedinger, 1989) was used to develop the T-year peak-flow regional regression equations for South Dakota.

Both dependent and explanatory variables were transformed to base-10 logarithms prior to all analyses to linearize the relation between dependent and explanatory variables and to normalize the distributions of residual errors. An iterative process was used to define subregional boundaries and select explanatory variables for the final GLS regression models. Initially, four subregions corresponding to those used for the generalized skew analysis were selected. Candidate variables for inclusion in the GLS regression models for these subregions were determined based on: plots of T-year peak flows versus basin and climatic characteristics; correlations of T-year peak flows with basin and climatic characteristics; all-possible-subsets OLS regressions of T-year peak flows as dependent variables versus basin and climatic characteristics as explanatory variables; and individual OLS regressions performed using the most influential explanatory variables to select a best-fit OLS regression model. Selection of the best-fit OLS regression models was based on (1) minimizing Mallow's C_p and the PRESS statistic (Helsel and Hirsch, 1992); and (2) passing of diagnostic checks to test for outliers, high-influence values, and multicollinearity between explanatory variables. The best-fit OLS regression model sets were then run using GLS regression. Residual errors of the GLS regressions were examined for spatial variability to evaluate whether subregional boundaries had been appropriately defined. Analysis of covariance (ANCOVA) was also performed to test for significant differences in the intercepts and explanatory variable coefficients between subregions. If the spatial variability in the residual errors and/or the ANCOVA indicated that subregional boundaries were inappropriately defined, the subregional boundaries were redefined to more appropriately fit the data; the OLS regression/GLS regression sequence was repeated until no further improvements in the GLS regression models could be made. Final GLS regression models also were selected based on (1) minimizing the model and sampling errors; (2) minimizing PRESS/*n*, an estimate of the mean prediction error sum of squares that indicates model performance when estimating peak-flow frequencies for ungaged sites (Gilroy and Tasker, 1990); (3) hydrologic reasonableness of the selected explanatory variables and the signs and magnitudes of their coefficients; and (4) physiographic and hydrologic reasonableness of the subregional boundary definitions. After multiple iterations of redefining subregional boundaries, performing the OLS/GLS regression sequence, and evaluating the residual errors, seven hydrologic subregions in South Dakota were defined (table 3 and fig. 3). A large-scale map (pl. 1) showing subregional boundaries, and locations of towns, major roads, county boundaries, and major streams is presented at the end of this report to aid in properly determining the subregion associated with a given basin.

Table 3. Descriptions of hydrologic subregions determined for the regional peak-flow magnitude and frequency analysis for South Dakota

Subregion	Description
A	Minnesota-Red River Lowland, Coteau des Prairies, and eastern part of the Southern Plateaus physical divisions of Flint (1955).
B	Lake Dakota Plain, James River Lowland and Highlands, and Coteau du Missouri physical divisions of Flint (1955); part of the Coteau du Missouri in central South Dakota that has topography typical of Great Plains "breaks" sites was excluded from this subregion.
C	Great Plains physiographic division of Fenneman (1946), excluding the Sand Hills influenced area in south-central South Dakota, and areas with topography typical of "breaks" sites, primarily in the Cheyenne, Bad, and White River basins.
D	Includes areas in the Great Plains physiographic division of Fenneman (1946) with topography typical of "breaks" sites.
E	Generally corresponds to the Sand Hills physical division of Flint (1955).
F	Generally corresponds to the northeast exterior part of the Black Hills physical division of Flint (1955).
G	Generally corresponds to the southwest interior part of the Black Hills physical division of Flint (1955).

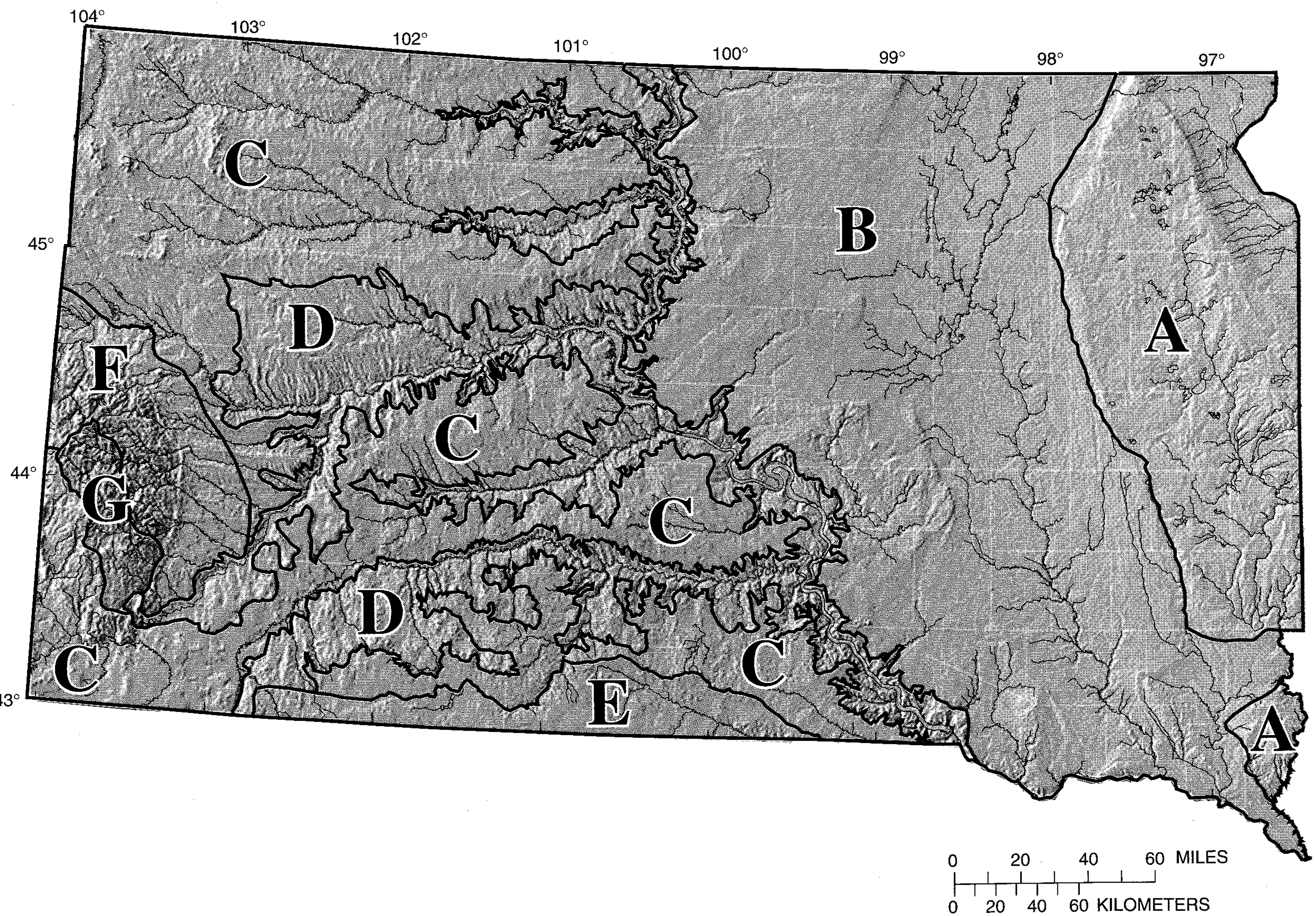


Figure 3. Hydrologic subregions determined for the regional peak-flow frequency analysis for South Dakota.

Subregion D is interspersed throughout subregion C. In developing the regression equations, it quickly became apparent that there were several sites located throughout the Great Plains of western South Dakota that yielded exceptionally large residuals when grouped with other sites in the Great Plains division. Examination of a shaded relief map revealed that these sites generally were located in areas of sharp relief, typically referred to as "breaks," and generally had larger main-channel slopes than most of the Great Plains sites and sites in subregion B (fig. 4). Therefore, a separate subregion (D) was developed primarily based on identifying the sharp relief areas in the Great Plains that were apparent on a shaded relief map. However, the distinction between subregions C and D is not discrete, as evidenced by the overlap in the statistical distributions of slopes (fig. 4). Some areas defined within subregion C appear somewhat similar topographically to areas of subregion D, especially in the upstream most part of the Cheyenne River Basin in Custer and Fall River Counties, and in the downstream part of the White River Basin east of the Little White River channel and south of the White River channel. Some of the sites in these areas appear to drain topography with moderate to fairly high relief, and yet they did not exhibit the same peak-flow frequency characteristics as the sites identified in subregion D. This may be due to local geologic influences or unrepresentative periods of record at these sites. The fact that the gaging station network is not very dense in the Bad and White River Basins also made it difficult to confidently define the boundary between subregions C and D in some areas. When estimating peak-flow frequencies in subregions C and D, users of the regression equations should pay particular attention to specific characteristics of the ungaged site relative to the characteristics for the gaging stations used to develop the regression equations for subregions C and D. Although the ranges in main-channel slopes for gaging stations in subregions C and D overlap (fig. 4), sites with main-channel slopes greater than about 60 ft/mi (feet per mile) generally are much more commonly found in subregion D than in subregion C. Conversely, sites with main-channel slopes less than about 60 ft/mi are much more commonly found in subregion C than in subregion D. Contributing drainage areas for gaging stations in subregion D tend to be small, with 75 percent of the stations having drainage areas less than 15 mi^2 (fig. 4). Although the largest contributing drainage areas for stations in subregion D exceed 100 mi^2, users should use extreme caution when applying the equations for subregion D to contributing drainage areas greater than 15 mi^2 because unusually large peak-flow estimates will result. Users should carefully examine the topographic characteristics of a given ungaged site, and the basin characteristics and peak-flow frequency relations for nearby or topographically similar gaging stations when selecting appropriate regression equations for a specific ungaged location in subregion C or D.

Regression equations for recurrence intervals of 2, 5, 10, 25, 50, 100, and 500 years were developed for hydrologic subregions A through G (table 4). The regional regression equations and the associated standard errors of estimate, standard errors of prediction, and equivalent years of record for the equations are listed in table 4. The standard error of estimate is an estimate of the square root of the mean-squared error in the GLS regression model that cannot be changed by collecting additional data. A procedure described by Hardison (1971) was used to convert the standard error of estimate to percent. The standard errors of estimate of the regression equations are a measure of the fit of the observed data to the regression model with the effects of time-sampling error in the observed values removed. They do not reflect the actual prediction errors made when using the models to predict T-year peak flows at ungaged sites. One must include the effect of estimating the regression coefficients from sample data (sampling mean-squared error) in order to estimate the standard error of a prediction. The average standard error of prediction is the square root of the sum of the average sampling error variance and the average model error variance. It is a measure of the average accuracy with which the regression model can estimate the T-year peak flow at an ungaged site.

The average equivalent years of record indicates the average number of years of streamflow record that provides an estimate equal in accuracy to the average standard error of prediction. The average standard error of prediction and associated average equivalent years of record reflect the prediction capabilities of the regression equations averaged for all of the sites used to develop the equations. The standard error of prediction for a specific site may vary considerably from the average standard error of prediction based on the basin characteristics of the specific site relative to the overall distribution of basin characteristics for all sites used to develop the equations. The variance-covariance matrix for the regression parameter estimates can be used to

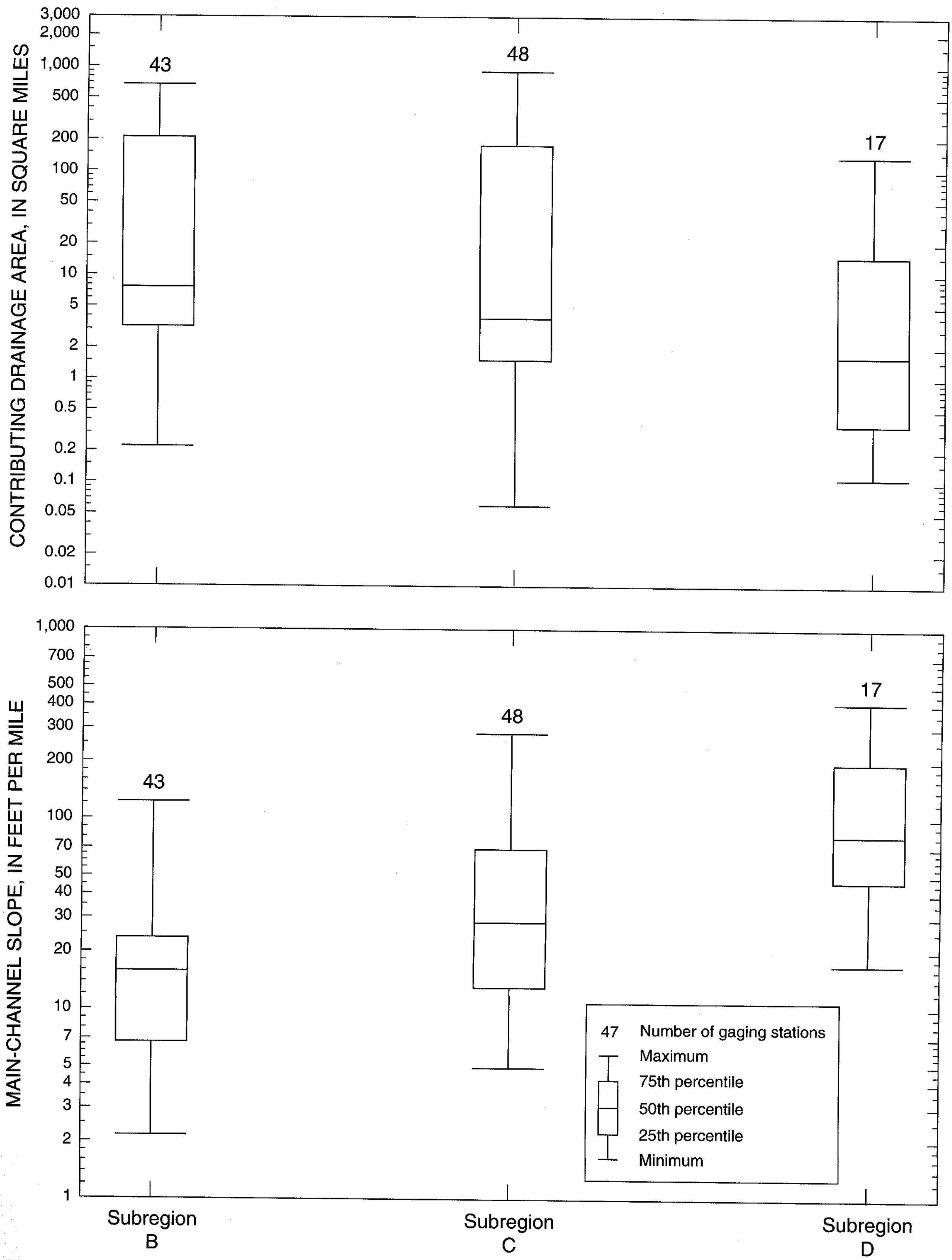


Figure 4. Statistical distributions of contributing drainage area and main-channel slope for gaging stations in subregions B, C, and D.

Table 4. Regional regression equations for South Dakota that relate peak flow magnitude for selected recurrence intervals to selected basin and climatic characteristics

[*CA*, contributing drainage area, in square miles; *S*, main-channel slope, in feet per mile; *PII*, precipitation intensity index]

Recurrence interval (years)	Equation	Number of stations used in analysis	Standard error of estimate (percent)	Average standard error of prediction (percent)	Average equivalent years of record (years)
		Subregion A			
2	$Q = 30.9\ CA^{0.513}\ PII^{6.14}$	55	55	59	4.5
5	$Q = 85.5\ CA^{0.509}\ PII^{5.45}$	55	50	54	6.1
10	$Q = 137\ CA^{0.510}\ PII^{5.12}$	55	50	54	7.8
25	$Q = 218\ CA^{0.513}\ PII^{4.80}$	55	51	56	9.8
50	$Q = 287\ CA^{0.517}\ PII^{4.62}$	55	53	58	11.0
100	$Q = 362\ CA^{0.521}\ PII^{4.47}$	55	55	61	11.9
500	$Q = 553\ CA^{0.531}\ PII^{4.22}$	55	62	69	13.0
		Subregion B			
2	$Q = 18.6\ CA^{0.425}\ PII^{1.10}$	43	60	67	5.4
5	$Q = 51.6\ CA^{0.508}\ PII^{0.835}$	43	57	64	7.1
10	$Q = 86.8\ CA^{0.546}\ PII^{0.764}$	43	59	67	8.7
25	$Q = 148\ CA^{0.584}\ PII^{0.730}$	43	62	72	10.6
50	$Q = 206\ CA^{0.606}\ PII^{0.728}$	43	65	76	11.6
100	$Q = 275\ CA^{0.625}\ PII^{0.742}$	43	69	81	12.4
500	$Q = 480\ CA^{0.661}\ PII^{0.811}$	43	78	93	13.6
		Subregion C			
2	$Q = 25.0\ CA^{0.569}$	48	104	108	1.8
5	$Q = 72.5\ CA^{0.578}$	48	65	67	4.8
10	$Q = 125\ CA^{0.579}$	48	55	58	8.3
25	$Q = 207\ CA^{0.573}$	46	50	53	12.0
50	$Q = 286\ CA^{0.570}$	46	50	53	14.9
100	$Q = 379\ CA^{0.566}$	46	51	55	16.5
500	$Q = 664\ CA^{0.556}$	46	61	65	16.6
		Subregion D			
2	$Q = 78.5\ CA^{0.357}$	17	98	109	2.3
5	$Q = 230\ CA^{0.455}$	17	54	61	7.4
10	$Q = 395\ CA^{0.515}$	17	37	44	17.9
25	$Q = 676\ CA^{0.585}$	17	26	34	39.1
50	$Q = 944\ CA^{0.627}$	17	22	33	52.5
100	$Q = 1{,}270\ CA^{0.663}$	17	22	34	59.2
500	$Q = 2{,}300\ CA^{0.732}$	17	27	41	57.5

Table 4. Regional regression equations for South Dakota that relate peak flow magnitude for selected recurrence intervals to selected basin and climatic characteristics—Continued

[*CA*, contributing drainage area, in square miles; *S*, main-channel slope, in feet per mile; *PII*, precipitation intensity index]

Recurrence interval (years)	Equation	Number of stations used in analysis	Standard error of estimate (percent)	Average standard error of prediction (percent)	Average equivalent years of record (years)
		Subregion E			
2	$Q = 12.1\ CA^{0.555}$	10	38	44	4.3
5	$Q = 18.9\ CA^{0.611}$	10	23	28	16.0
10	$Q = 22.6\ CA^{0.653}$	10	20	26	27.0
25	$Q = 27.0\ CA^{0.702}$	10	23	30	30.2
50	$Q = 30.3\ CA^{0.737}$	10	28	36	27.4
100	$Q = 33.6\ CA^{0.769}$	10	34	42	24.2
500	$Q = 41.4\ CA^{0.840}$	10	49	60	18.5
		Subregion F			
2	$Q = 0.937\ CA^{0.676}\ S^{0.447}$	17	93	107	2.6
5	$Q = 0.591\ CA^{0.779}\ S^{0.745}$	17	71	83	6.0
10	$Q = 0.471\ CA^{0.832}\ S^{0.907}$	17	61	73	10.5
25	$Q = 0.406\ CA^{0.888}\ S^{1.06}$	17	53	66	18.4
50	$Q = 0.381\ CA^{0.925}\ S^{1.16}$	17	50	64	24.6
100	$Q = 0.352\ CA^{0.960}\ S^{1.25}$	17	49	64	29.4
500	$Q = 0.243\ CA^{1.04}\ S^{1.47}$	17	58	78	31.2
		Subregion G			
2	$Q = 3.46\ CA^{0.650}$	7	41	51	3.9
5	$Q = 7.70\ CA^{0.654}$	7	58	71	3.2
10	$Q = 11.3\ CA^{0.673}$	7	70	87	3.2
25	$Q = 16.5\ CA^{0.704}$	7	86	108	3.3
50	$Q = 21.0\ CA^{0.731}$	7	98	126	3.3
100	$Q = 25.8\ CA^{0.759}$	7	110	144	3.4
500	$Q = 38.5\ CA^{0.826}$	7	141	193	3.5

estimate the standard error of prediction, prediction intervals, and equivalent years of record for a specific ungaged site (Hodge and Tasker, 1995). The variance-covariance matrix for the regression parameters of the equations in table 4 are presented in table 7 in the Supplemental Information section. A computer program that calculates the predicted values for each quantile, the standard error of prediction, prediction intervals, and equivalent years of record for a specific ungaged site is available upon request from the USGS in Rapid City, South Dakota.

For subregions A and B, the explanatory variables are contributing drainage area and precipitation intensity index, which is defined as the 2-year, 24-hour precipitation intensity (U.S. Weather Bureau, 1961), in inches, minus 1.5. A map showing isolines for the precipitation intensity index is presented in figure 5. The regression equations were developed using precipitation intensity index values from figure 5, and therefore, only values from figure 5 should be used for input to the regression equations. More recently developed local precipitation intensity data would not be appropriate for use in the regression equations.

For subregions C, D, E, and G, contributing drainage area was the only significant explanatory variable in the regression equations. For subregion F, contributing drainage area and main-channel slope were the explanatory variables.

Standard errors of estimate for subregion G tended to be much higher than the other subregions, especially for recurrence intervals greater than 25 years. Generally, peak flows in the Black Hills are highly variable and difficult to regionalize. The delineation of subregions F and G was based on the general observation that peaks in the interior part of the Black Hills generally were not as large or variable as those in the exterior part where high-intensity spring and summer storms are more common. The large standard errors of estimate in subregion G, which are in part due to the small number of sites (7), indicate that variation in peak flows throughout the Black Hills is too complex to be effectively described by the delineation of two subregions. Other definitions of subregional boundaries in the Black Hills were investigated but did not improve model results. Also, application of design probability theory (Riggs, 1968) was investigated as an alternative approach for regionalizing peak-flow magnitudes and frequencies in the Black Hills interior area, but the large variability in peak-flow characteristics also yielded substantial uncertainties in the results of that approach. Users of the regression equations for subregion G should be aware of the large standard errors of estimate for larger recurrence interval equations and of the fact that these equations were developed using data from only seven gaging stations. Accordingly, peak-flow magnitude and frequency data from nearby gaging stations should be examined and used as a guide to determine whether the use of regression equations for subregion G is appropriate.

Estimation of peak-flow magnitudes and frequencies for streams draining the Black Hills is further complicated by the fact that some of those streams cross outcrops of fractured limestone bedrock. Large losses in streamflow can occur where streams cross these outcrops. Caution should be used in applying the regression equations for subregions F and G to ungaged locations that are immediately downstream from limestone bedrock outcrops. This is especially true for estimates of peak flows with smaller recurrence intervals.

The regression equations developed for subregions A through G may be used to determine peak flows for ungaged sites on ungaged streams. Examples of the use of the regional regression equations to compute peak-flow values for ungaged sites on ungaged streams are given in the section Examples of Estimating Peak-Flow Magnitudes for Selected Frequencies for Ungaged Sites and for Gaging Stations.

Limitations on Use of the Regression Equations

The following limitations should be considered when using the regression equations to compute peak-flow frequencies for South Dakota streams: (1) The equations apply to streams that are located in rural watersheds and should not be applied to watersheds substantially affected by urbanization (peak flows at sites substantially affected by urbanization can be estimated using techniques presented in Sauer and others, 1983); (2) the equations should not be used where dams, flood-detention structures, and other manmade works exist that significantly affect the annual peak flows; and (3) the equations generally should be used only for streams that have drainage areas of less than or equal to 1,000 mi^2 and for streams that have drainage areas and basin and climatic characteristics that are within the range of characteristics used to develop the regression equations. The ranges of the characteristics for the gaging stations used to develop the regression equations are given in table 5. Although gaging stations with contributing drainage areas up to 137 mi^2 were used to develop equations for subregion D, much caution should be used when applying these equations to basins with contributing drainage areas greater than 15 mi^2.

Table 5. Ranges of basin and climatic characteristics used to develop the regional regression equations

[--, not applicable]

Subregion	Contributing drainage area (square miles) (*CA*)	Precipitation intensity index (*PII*)	Main-channel slope (feet per mile) (*S*)
A	0.14 - 983	0.79 - 1.30	--
B	0.22 - 670	0.60 - 1.21	--
C	0.06 - 904	--	--
D	0.11 - 137	--	--
E	10.0 - 760	--	--
F	0.63 - 920	--	29.6 - 460
G	3.81 - 105	--	--

WEIGHTED PEAK-FLOW MAGNITUDE AND FREQUENCY RELATIONS FOR GAGING STATIONS

A weighting procedure can be used to combine peak-flow magnitude and frequency information from the regional regression equations with magnitude and frequency information from the systematic record for gaging stations. Weighted peak-flow magnitudes for various frequencies were calculated for the 197 gaging stations used in the regional analysis using the following equation developed from principles discussed in Bulletin 17B (G.D. Tasker, written commun., 1997):

$$Q_{TW} = \frac{n\,Q_{TS} + en\,Q_{TR}}{n + en} \tag{4}$$

where

Q_{TW} = weighted peak flow, in cubic feet per second, for recurrence interval of T years;

n = number of years of station data used to compute Q_{TS};

Q_{TS} = station peak flow, in cubic feet per second, for recurrence interval of T years;

en = equivalent years of record for Q_{TR}; and

Q_{TR} = peak flow, in cubic feet per second, for recurrence interval of T years from regional regression equation (as discussed in the section Development of Regression Equations) that relates peak flow to basin characteristics.

The equivalent years of record, *en*, is a measure of the accuracy of prediction in terms of the number of years of record that is required for each gaging station to achieve results of equal accuracy to that of the regional regression equation. A further explanation on how the equivalent years of record is calculated is given by Hardison (1971). Weighted peak-flow magnitudes and frequencies for the 197 gaging stations used in the regional analysis are included in table 6. The weighted peak-flow magnitudes generally should provide better estimates than the station peak-flow magnitudes because they take into account additional regional information developed from the regional analysis and reduce the time-sampling error associated with short periods of record.

PEAK-FLOW MAGNITUDE AND FREQUENCY RELATIONS FOR UNGAGED SITES NEAR A GAGING STATION ON THE SAME STREAM

The following equation can be used to determine T-year peak-flow values for an ungaged site located near a gaging station on the same stream. Generally, this equation should be used when the contributing drainage area for the ungaged site is from 75 to 150 percent of the contributing drainage area for the gaged site (otherwise, the regional regression equations discussed earlier in this report should be used):

$$Q_{T(u)} = Q_{TW(g)}\,(CA_u/CA_g)^x \tag{5}$$

where

$Q_{T(u)}$ = peak flow, in cubic feet per second, for the ungaged site for a recurrence interval of T years;

$Q_{TW(g)}$ = weighted peak flow, in cubic feet per second, for the gaging station for a recurrence interval of T years;

CA_u = contributing drainage area, in square miles, for the ungaged site;

CA_g = contributing drainage area, in square miles, for the gaging station;

x = mean exponent for the appropriate hydrologic region; for subregion A, the mean exponent is 0.529; B, 0.615; C, 0.569; D, 0.545; E, 0.691; F, 0.654; and G, 0.689.

The peak flow, $Q_{TW(g)}$, is the weighted peak-flow value in table 6 that was determined using procedures discussed in the section Weighted Peak-Flow Magnitude and Frequency Relations for Gaging Stations. The mean exponent, x, was determined by regressing the dependent variable (logarithm of T-year peak flow) on the independent variable (logarithm of contributing drainage area) for 2-, 5-, 10-, 25-, 50-, 100-, and 500-year recurrence intervals. The mean exponent for the seven recurrence intervals was then computed for each hydrologic subregion. Values of the mean exponent (x) for each subregion are presented above.

EXAMPLES OF ESTIMATING PEAK-FLOW MAGNITUDES FOR SELECTED FREQUENCIES FOR UNGAGED SITES AND FOR GAGING STATIONS

Examples are presented for estimating peak-flow magnitudes for selected frequencies for ungaged sites and for gaging stations. The examples presented include (1) using the regional regression equations for ungaged sites on ungaged streams; (2) computing weighted peak-flow magnitudes for gaging stations; (3) computing peak-flow magnitudes for an ungaged site near a gaging station on the same stream; and (4) computing peak-flow magnitudes for an ungaged site between two gaging stations on the same stream.

Regression Equations to Compute Peak-Flow Magnitudes for Selected Frequencies for Ungaged Sites on Ungaged Streams

Depending on the location of the ungaged site and its relation to the hydrologic subregion boundaries, one of two procedures can be used to compute peak-flow values. Procedure 1 should be used when the stream drainage area is within a single subregion. Procedure 2 should be used when the stream drainage area is part of more than one subregion.

Use procedure 1 to determine the 100-year peak flow for an ungaged site on Mosquito Creek near Marty, SD. The site is located in subregion B (fig. 3). The basin characteristics that were determined from a USGS 7.5-minute topographic map and figure 5 are

CA = 9.2 mi^2, and
PII = 0.98 (fig. 5).

The indicated peak flow for a recurrence interval of 100 years, based on the appropriate equation from table 4, is

$$Q_{100} = 275\,(9.2)^{0.625}\,(0.98)^{0.742} = 1,080 \text{ ft}^3/\text{s}.$$

Procedure 2 is similar to procedure 1 except T-year regional regression equations would be solved for each of the associated subregions and the results would be averaged or apportioned according to the fraction of the contributing drainage area that is in each subregion. Use procedure 2 to determine the 100-year peak flow for an ungaged site on Willow Creek at Highway 34 about 7 mi west of Pierre, S. Dak. The site is located partly in subregion C and partly in subregion D (fig. 3). The basin characteristics that were determined from a USGS 7.5-minute topographic map are

CA = 86.5 mi^2,
Contributing drainage area in subregion C = 66.9 mi^2, and
Contributing drainage area in subregion D = 19.6 mi^2.

The total contributing drainage area is used to calculate peak flow for a recurrence interval of 100 years using the 100-year regression equations for both subregion C and D. The results of those calculations are then averaged by weighting according to the proportion of the drainage area in each subregion.

The indicated peak flow for a recurrence interval of 100 years for subregion C is

$$Q_{100} = 379\,(86.5)^{0.566} = 4,730 \text{ ft}^3/\text{s}.$$

The indicated peak flow for a recurrence interval of 100 years for subregion D is

$$Q_{100} = 1,270\,(86.5)^{0.663} = 24,400 \text{ ft}^3/\text{s}.$$

The results of the two equations are then averaged according to the proportion of the drainage area in each subregion:

$$Q_{100} = 4,730\left(\frac{66.9}{86.5}\right) + 24,400\left(\frac{19.6}{86.5}\right) = 9,190 \text{ ft}^3/\text{s}.$$

Weighted Peak-Flow Magnitudes for Selected Frequencies for Gaging Stations

The procedure used to determine the weighted peak flow (eq. 4) for the 100-year recurrence interval for station 06477400, Firesteel Creek tributary near Wessington Springs, SD (table 6, site 144), is shown in the following example.

Station 06477400:

n = 12 years,

Q_{TS} = 92 ft^3/s,

CA = 0.22 mi^2, and

PII = 0.91 (from fig. 5).

The gaging station is located in subregion B, and the 100-year equation from table 4 is

$$Q_{100} = 275\ CA^{0.625} PII^{0.742}; \quad \text{thus,}$$

$$Q_{TR} = Q_{100} = 275\,(0.22)^{0.625}\,(0.91)^{0.742} = 100\ \text{ft}^3/\text{s}.$$

The average equivalent years of record, *en*, for subregion B for the 100-year equation is 12.4 years (table 4). Using equation 4,

$$Q_{TW} = (12\,(92) + 12.4\,(100))/(12 + 12.4) = 96\ \text{ft}^3/\text{s}.$$

Peak-Flow Magnitudes for Selected Frequencies for an Ungaged Site near a Gaging Station on the Same Stream

The following is an example of how to use the drainage-area ratio method (eq. 5) to determine peak flow for an ungaged site near a gaging station on the same stream. To estimate the 100-year peak flow for an ungaged site on Oak Creek near Mahto, locate the site on figure 3. The site is located upstream from a gaging station on the same stream (station 06354882, Oak Creek near Wakpala, SD, site 29). Contributing drainage area (CA_u) for the ungaged site is 269 mi^2, and the contributing drainage area (CA_g) for the gaging station (table 6, site 29) is 356 mi^2. Both the ungaged site and the gaging station are located in subregion C (fig. 3). Determine if the drainage-area ratio (CA_u/CA_g) is between 0.75 and 1.5:

$$CA_u/CA_g = 269/356 = 0.76,$$

which meets the drainage-area ratio guideline. Thus, the following relation (eq. 5) is used:

$$Q_{100\,(u)} = Q_{100\,(g)}\,(CA_u/CA_g)^x$$

where

$Q_{100(g)}$ = 10,300 ft^3/s, the weighted peak flow for the gaging station (table 6);

CA_u = 269 mi^2;

CA_g = 356 mi^2; and

x = 0.569 (from page 21).

Therefore,

$$Q_{100\,(u)} = 10,300\,(269/356)^{0.569} = 8,780\ \text{ft}^3/\text{s}.$$

Peak-Flow Magnitudes for Selected Frequencies for an Ungaged Site Between Two Gaging Stations on the Same Stream

An ungaged site for which a peak-flow calculation is desired sometimes may be between two gaging stations on the same stream. If the contributing drainage area for the ungaged site is within 75 to 150 percent of the contributing drainage area for both of the gaging stations, the drainage-area ratio method (eq. 5) should be applied to determine peak flow for the ungaged site ($Q_{T(u)}$) using data from each of the gaging stations. The resulting two peak-flow values then are averaged in logarithm units. The peak flow ($Q_{T(u)}$) also can be determined by interpolating in logarithm units between the gaging stations.

The following examples illustrate how to use equation 5 and interpolation to determine a 100-year peak flow for an ungaged site that is between two gaging stations on the same stream. The gaging stations are station 06447500, Little White River near Martin, SD (table 6, site 111), and station 06449100, Little White River near Vetal, SD (table 6, site 113). Contributing drainage area (CA_u) for the ungaged site is 325 mi^2, contributing drainage area (CA_g) for gaging station 06447500 is 230 mi^2, and contributing drainage area (CA_g) for gaging station 06449100 is 415 mi^2. The drainage basins for the ungaged site and the gaging stations mostly are located in subregion E (fig. 3).

Determine if the drainage-area ratio (CA_u/CA_g) for the ungaged site relative to each gaging station is between 0.75 and 1.50:

1. Station 06447500:
   $CA_g = 230\ mi^2$,
   $CA_u = 325\ mi^2$,
   $CA_u/CA_g = 325/230 = 1.41$.

2. Station 06449100:
   $CA_g = 415\ mi^2$,
   $CA_u = 325\ mi^2$,
   $CA_u/CA_g = 325/415 = 0.78$.

The drainage-area ratio guideline has been met for both situations. Thus equation 5 is applied for the ungaged site using data from both gaging stations, and the results are averaged in logarithm units.

1. Station 06447500:
   $Q_{100(g)} = 2{,}110\ ft^3/s$, weighted peak flow for the gaging station (table 6),
   $Q_{100(u)} = 2{,}110\ (325/230)^{0.691} = 2{,}680\ ft^3/s$.

2. Station 06449100
   $Q_{100(g)} = 3{,}500\ ft^3/s$
   $Q_{100(u)} = 3{,}500\ (325/415)^{0.691} = 2{,}960\ ft^3/s$.

The logarithmic average is

$$\frac{[\log(2,680) + \log(2,960)]}{2} = \frac{(3.428 + 3.471)}{2} = 3.450$$

$$Q_{100(u)} = \text{antilog}\ (3.450) = 2,820\ ft^3/s\ .$$

The peak flow for the ungaged site determined by interpolation between the two gaging stations is shown below:

$$\log Q_{100(u)} = \log Q_{100\,(06449100)} - \left[\frac{\log CA_{06449100} - \log CA_u}{\log CA_{06449100} - \log CA_{06447500}}\right] \text{x}$$
$$[\log Q_{100\,(06449100)} - \log Q_{100\,(06447500)}]$$

$$\log Q_{100(u)} = \log(3,500) - \left[\frac{\log(415) - \log(325)}{\log(415) - \log(230)}\right] \text{x}$$
$$[\log(3,500) - \log(2,110)]$$

$$\log Q_{100(u)} = 3.544 - \left[\frac{2.618 - 2.512}{2.618 - 2.362}\right](3.544 - 3.324) = 3.453$$

$$Q_{100(u)} = \text{antilog}\ (3.453) = 2,840\ ft^3/s.$$

DATA NEEDS FOR IMPROVING REGIONAL PEAK-FLOW FREQUENCY ANALYSES

Future peak-flow frequency studies for South Dakota could be improved if additional gaging stations were established or if previously discontinued gaging stations were re-established on natural-flow streams, especially for sites on streams with drainage areas of less than 100 mi^2. Most of the gaging stations operated during 1994 are on streams with contributing drainage areas greater than 100 mi^2 and their primary purpose is not to define natural variations in hydrologic characteristics. In addition, determination of basin and climatic characteristics using GIS coverages may be useful.

A description of the history of the streamflow-data program for South Dakota appears in Larimer (1970). During 1956-80, many crest-stage gages were operated to define peak-flow frequencies on streams with small drainage areas (Becker, 1974, 1980). However, following the completion of Becker's studies, the number of gaging stations in smaller basins has decreased. In Burr and Korkow (1996), peak-flow frequencies were reported for 234 gaging stations; however, 105 of those stations were still in operation in 1994. Of these stations, many were on regulated streams or on streams with drainage areas greater that 1,000 mi^2. Thus, only about 50 of the stations still in operation in 1994 were appropriate for developing regional regression equations for use in estimating peak-flow values for unregulated streams, and only 10 of those gaging stations were located in basins with drainage areas less than 10 mi^2. The emphasis in the streamflow-gaging network has moved away from collecting data to define regional variation in natural hydrologic characteristics and instead focuses on providing streamflow information at sites that are of particular interest to the agencies that cooperatively participate in funding the stations. Operation of additional stations would help to adequately define regional variation in natural hydrologic characteristics. There is particular importance for additional stations in (1) smaller basins (less than 100 mi^2) throughout the State; (2) subbasins in the Bad and White River Basins to more confidently define boundaries between subregions C and D; and (3) basins in the southwest corner of the State to more confidently define boundaries between subregions C, D, F, and G. The pertinent streamflow-gaging data could be obtained by operating continuous-record gaging stations and (or) crest-stage gages for at least 25 years, an adequate record length for estimating the 100-year peak flow. The GLS procedure could be used to perform a network analysis that would yield a relation between the effect of adding or deleting stations to an operating budget. The GLS procedure also would relate the standard error (at site or regional) to record length in order to reduce the standard error (at site or regional) to a certain level, and would be beneficial in determining specific gaging locations that would be most useful for a regional analysis.

Recently, there have been significant advances in development of GIS coverages with information on basin and climatic characteristics. This type of information may be readily available to most users of the regional regression equations. Use of GIS coverages may provide a more accurate and convenient method of estimating the basin and climatic characteristics that are used in the regional analysis.

SUMMARY

A generalized skew coefficient analysis was completed for South Dakota to test the validity of using the generalized skew coefficient map in Bulletin 17B. The generalized skew coefficients for South Dakota in Bulletin 17B were evaluated for applicability by comparing them to generalized skew coefficients determined by the three recommended methods: (1) lines of equal skew coefficient drawn on a Statewide map; (2) regression equations that relate skew coefficient to selected basin and climatic characteristics; and (3) mean values of skew coefficient within subregions of interest. Data for 106 gaging stations with 25 or more years of unregulated systematic record were used in the generalized skew analysis. Results of the analysis indicate that the Bulletin 17B generalized skew coefficient map generally provides adequate generalized skew coefficients for estimating peak-flow magnitudes and frequencies for South Dakota gaging stations.

Peak-flow records through 1994 for 197 continuous- and partial-record streamflow-gaging stations that had 10 or more years of unregulated systematic record were used in a generalized least-squares regression analysis that relates peak flows for selected recurrence intervals to selected basin and climatic characteristics. Peak-flow equations were developed for recurrence intervals of 2, 5, 10, 25, 50, 100, and 500 years for seven hydrologic subregions in South Dakota. The peak-flow equations are applicable to natural-flow streams that have drainage areas less than or equal to 1,000 mi^2. Basin and climatic characteristics in the final regression equations included contributing drainage area, main-channel slope, and precipitation intensity index. The standard error of estimate for the seven hydrologic subregions ranges from 22 to 110 percent for the 100-year peak-flow equations. Weighted peak flows for various frequencies based on gaging-station data and the regional regression equations are provided for each gaging station. The weighted peak flows generally should provide better peak-flow frequency estimates than the station peak flows because they include additional regional information developed from the regional analysis and reduce the time-sampling error associated with short periods of record.

Examples are given for (1) determining peak-flow magnitudes and frequencies for ungaged sites on ungaged streams using the regional regression equations; (2) determining weighted peak-flow magnitudes and frequencies for gaging stations; and (3) using the drainage-area ratio method for determining peak-flow magnitudes and frequencies for ungaged sites near a gaging station on the same stream and ungaged sites between two gaging stations on the same stream.

REFERENCES CITED

Becker, L.D., 1974, A method for estimating magnitude and frequency of floods in South Dakota: U.S. Geological Survey Water-Resources Investigations 35-74, 78 p. (available only from U.S. Department of Commerce, National Technical Information Service, Springfield, VA 22151 as report PB-239 831).

———1980, Techniques for estimating flood peaks, volumes, and hydrographs on small streams in South Dakota: U.S. Geological Survey Water-Resources Investigations Report 80-80, 82 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, 64 p.

Benson, M.A., and Carter, R.W., 1973, A national study of the streamflow data-collection program: U.S. Geological Survey Water-Supply Paper 2028, 44 p.

Benson, R.D., 1986, South Dakota surface-water resources, *in* National water summary 1985 -- Hydrologic events and surface-water resources: U.S. Geological Survey Water-Supply Paper 2300, p. 419-424.

Benson, R.D., Hoffman, E.B., and Wipf, V.J., 1985, Analyses of flood-flow frequency for selected gaging stations in South Dakota: U.S. Geological Survey Water-Resources Investigations Report 85-4217, 202 p.

Burr, M.J., and Korkow, K.L., 1996, Peak-flow frequency estimates through 1994 for gaged streams in South Dakota: U.S. Geological Survey Open-File Report 96-202, 407 p.

Dempster, G.R., 1983, Streamflow basin characteristics, Chapter II, Section A, in WATSTORE user's guide, v. 4: U.S. Geological Survey Open-File Report 79-1336-I, p. A-24 to A-34.

Fenneman, N.M., 1946, Physical divisions of the United States: U.S. Geological Survey map prepared in cooperation with the Physiographic Commission, U.S. Geological Survey, scale 1:700,000 (reprinted 1964).

Flint, R.F., 1955, Pleistocene geology of eastern South Dakota: U.S. Geological Survey Professional Paper 262, 173 p.

Gilroy, E.J., and Tasker, G.D., 1990, Multicollinearity and influential observations in hydrologic model selection, *in* Berk, K., and Malone, L. (ed.), Computing Science and Statistics: Proceedings of the 21st Symposium on the Interface, p. 350-354.

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

Helsel, D.R., and Hirsch, R.M., 1992, Statistical methods in water resources: Studies in Environmental Science 49, New York, Elsevier Science, 522 p.

Hodge, S.A., and Tasker, G.D., 1995, Magnitude and frequency of floods in Arkansas: U.S. Geological Survey Water-Resources Investigation Report 95-4224, 52 p.

Hoffman, E.B., Freese, M.E., and Winter, D.R., 1986, Analyses of flood-flow frequency for selected gaging stations in South Dakota through September 1985: U.S. Geological Survey Open-File Report 86-496, 137 p.

Larimer, O.J., 1970, A proposed streamflow data program for South Dakota: U.S. Geological Survey open-file report, 46 p.

McCabe, J.A., and Crosby, O.A., 1959, Floods in North and South Dakota, Frequency and magnitude: U.S. Geological Survey open-file report, 132 p.

Patterson, J.L., 1966, Magnitude and frequency of floods in the United States, pt. 6-A, Missouri River basin above Sioux City, Iowa: U.S. Geological Survey Water-Supply Paper 1679, 471 p.

Patterson, J.L., and Gamble C.R., 1968, Magnitude and frequency of floods in the United States, pt. 5, Hudson Bay and Upper Mississippi River Basins: U.S. Geological Survey Water-Supply Paper 1678, 546 p.

Riggs, H.C., 1968, Frequency curves: U.S. Geological Survey Techniques of Water-Resources Investigations, book 4, chap. A1, p. 13.

Sauer, V.B., Thomas, W.O., Stricker, V.A., and Wilson, K.V., 1983, Flood characteristics of urban watersheds in the United States: U.S. Geological Survey Water-Supply Paper 2207, 63 p.

Stedinger, J.R., and Tasker, G.D., 1985, Regional hydrologic analysis 1--Ordinary, weighted, and generalized least squares compared: Water Resources Research, v. 21, no. 9, p. 1421-1432.

Tasker, G.D., and Stedinger, J.R., 1986, Regional skew with weighted least squares regression: Journal of Water Resources Planning and Management, v. 112, no. 2, p. 225-237.

———1989, An operational GLS model for hydrologic regression: Journal of Hydrology, v. 111, p. 361-375.

U.S. Geological Survey, 1975, Mineral and water resources of South Dakota: U.S. 94th Congress, 1st Session, Interior and Insular Affairs Committee Print, Washington, D.C., 313 p.

U.S. Weather Bureau, 1959a, Climates of the States, South Dakota: U.S. Weather Bureau, Climatography of the United States, No. 60-39, 15 p.

———1959b, Evaporation maps for the United States: Technical Paper 37, 13 p.

———1961, Rainfall frequency atlas of the United States: Technical Paper 40, 115 p.

———1964, Frequency of maximum water equivalent of March snow cover in north central United States: Technical Paper 50, 24 p.

United States Water Resources Council, 1982, Guidelines for determining flood flow frequency: Hydrology Subcommittee, Bulletin 17B, appendices 1-14, 28 p.

Williams-Sether, Tara, 1992, Techniques for estimating peak-flow frequency relations for North Dakota streams: U.S. Geological Survey Water-Resources Investigations Report 92-4020, 57 p.